McGraw-Hill Encyclopedia of Science and Technology

McGraw-Hill Encyclopedia

McGRAW-HILL BOOK COMPANY INC.

NEW YORK CHICAGO SAN FRANCISCO DALLAS TORONTO LONDON



of Science and Technology

AN INTERNATIONAL REFERENCE WORK

IN FIFTEEN VOLUMES INCLUDING AN INDEX

VOLUME 13 SPIR-TOU



(LEFT) Photomicrograph of copper sulfate and calcium chloride crystals *(photograph by L. C. Massopust)*
(RIGHT) A model and its mirror image, of the reciprocal space of the bismuth lattice. The lines mark the intersection of principal planes in reciprocal space, while the inner solid represents the shape of the first Brillouin zone *(Franklin Institute)*.

Library of Congress Catalog Card Number 60-11000

Suggestions to the Readers

The basic plan of the Encyclopedia is explained here in order to facilitate its use

The subject matter of the various disciplines or branches of science and technology is organized systematically a general article provides a broad survey of the field, and a number of separate articles, alphabetically arranged cover its main subdivisions and more specific aspects

Cross references guide the reader from the general articles to the other articles into which the subject is subdivided, and from these to articles on more highly specialized phases of the subject The cross references—there are about 50,000 of them—are printed in small capital letters so that they can be easily recognized By means of the cross references a reader may find his way from ELECTRICAL ENGINEERING, through ELECTRONICS and VACUUM TUBE, to ELECTRON MOTION IN VACUUM or ELECTRON EMISSION. Or, following another line of cross references, the reader would be led to ELECTRIC POWER SYSTEMS, TRANSMISSION LINES, ELECTROMAGNETIC WAVE, and so on

In general each article begins with a definition of the title that states its scope and coverage Usually, only the scientific or technological sense is discussed. Most of the articles, after this statement, go on to increasingly complex and detailed considerations A reader thus needs to proceed only as far as his inclinations and requirements dictate.

The Index Volume 15, should be consulted to locate the discussion of topics covered in the Encyclopedia but not given in separate entries

Every phylum, class, and order in the plant and animal kingdoms is allotted a separate article Many of the more common families, genera, and species are covered either in one of the order articles or in a separate article under its own scientific or common name

The adjectives electric and electrical are used in the following senses Electric—containing, producing, arising from, actuated by, or carrying electricity, or capable of doing so; as, for instance, electric generator, electric motor, electric wiring. Electrical—related to, pertaining to, or associated with electricity, but not having its properties or characteristics; as, for example, electrical code, electrical engineering

Words used as titles are, wherever possible, given in the singular to permit a consistent alphabetic arrangement. Titles are alphabetized by word and not by letter; for example,

Earth sciences
Earth tides
Earthmover
Earthquake

A word used as a noun precedes the same word used adjectivally, thus

Mercury (element)
Mercury (planet)
Mercury battery

or

Circuit, electronic
Circuit breaker

Hyphened terms are alphabetized as single words, for example,

Animal virus
Animal-feed composition

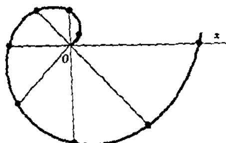
Most of the longer articles contain bibliographies citing useful sources of further information For additional bibliographical citations, the reader should refer to related articles (as indicated by the cross references in the article) Bibliographies are placed at the ends of articles or sometimes at the ends of major sections in long articles

A list of initials and names of the contributors to the Encyclopedia is to be found in Volume 15. This list will permit quick identification of a contributor's initials after an article Immediately following this list is a second list of encyclopedia contributors with their affiliations and the titles of articles each has written for the Encyclopedia

McGraw-Hill Encyclopedia of Science and Technology

Spiral

A term used generically to describe any geometrical entity that winds about a central point or axis while also receding from it. Spiral staircases, helices, nonplanar loxodromes (curves that intersect those of a given class at a constant angle, for example, rhumb lines, in case the curves are on a sphere whose meridians form the given class) are



Planar spiral curve (Archimedes' spiral).

examples of spirals whose windings do not lie in a plane. Archimedes introduced the planar spiral curve $\rho = a\theta$, called by his name, in his book *On Spirals*. Other important planar spiral curves are logarithmic, $\rho = e^{a\theta}$, the tangents to which make a constant angle with the radii vectors drawn to the points of contact, hyperbolic, $\rho\theta = a$, and the lituus, $\rho^2\theta = a^2$. *See* HELIX. [L.M.BL.]

Spirillaceae

A fam...

...blooded animals. Species cause diseases in man like cholera and rat bite fever, and abortion in sheep. The Pseudomonadales comprise an artificial grouping of chemoautotrophic, gram-negative, curved to spirally twisted rods, with rigid cell walls and polar flagella. The family is subdivided into genera on the ...
number an...
lism. A de...
column in t... ...ditions aerobic bacteria require oxygen and grow best at high oxygen tension of 20% or more, anaerobic bacteria do not require oxygen for growth, facultative anaerobic bacteria grow under aerobic or anaerobic conditions, heterotrophic bacteria are unable to grow with carbon dioxide as a sole carbon source and require organic compounds, autotrophic bacteria utilize only inorganic materials as a source of nutrients and carbon dioxide as a sole source of carbon. As in other areas of bacterial taxonomy, authorities differ widely both on the species to be included in the family and its subdivision. *See* PSEUDOMONADALES.

Morphology. Cell curvature gives a unique and easily recognizable form to the typical *Spirillaceae*. The vibrios, having a single bend, are comma-shaped to crescent shaped organisms. The spirilla have two or more bends, twists, or both, resulting in S-shaped, wavelike, or spirally coiled cells. The dividing lines between the straight rods, or pseudomonads, the vibrios, and the spirilla are indistinct, and no absolute separation can be made on a morphological basis. The rigidity of the spirilla distinguishes them from the spirochetes (see SPIROCHAETALES). However, the spirilla, especially the longer forms, not only can bend or flex, and thus distort the basic wave pattern, but also elongate during motion. Thus cell rigidity is a relative matter, and the existence of forms with intermediate properties like *Spirillum minus* causes taxonomic difficulties. *See* RAT-BITE FEVER.

The fundamental basis for a curved cell structure has not been investigated. The curvature of the cell is influenced by environmental factors. For example, under different conditions of culture, actively growing *Spirillum volutans*, the largest of the spirilla, can occur as almost straight rods, as evenly bent S-shaped cells, as angular, tightly coiled forms, or as closed rings. The typical curved form can be lost permanently, yielding straight rods on laboratory culture.

Certain spirilla have a life cycle; they exist in morphologically distinct vegetative and resting stages. The spiral cells at the end of active growth gradually shorten and round out, forming oval or spherical microcysts. These, in a favorable environment, germinate by unipolar or bipolar emergence of the vegetative form. The formation of ringlike bodies, reported for the genus *Microcyclus*, may be part of a similar life cycle.

Metabolism. Most of the *Spirillaceae* that have been cultured are nonexacting nutritionally and are oxidative ...
obic
obtai...
sulfa... ...carbon dioxide, respectively, replacing oxygen as the terminal hydrogen acceptors. Cytochrome pigments have been demonstrated in *Desulfovibrio*. The fermentation of sugars and amino

Genera of the family Spirillaceae*

Genus	Characteristic cell shape	Flagella, no and position	Metabolism
Vibrio	Single curve, comma-shaped	1, polar	Aerobic to anaerobic, heterotrophic
Desulfovibrio	Single curve, comma-shaped	1, polar	Anaerobic, heterotrophic [illegible]
Methanobacterium	Straight to slightly curved	None	[illegible] trophic, carbon dioxide reduced to methane
Cellvibrio	Slightly curved long rods	1, polar	Aerobic to facultatively anaerobic, heterotrophic, oxidizes cellulose
Cellfalcicula	Straight to curved, spindle-shaped	1, polar	Like *Cellvibrio*
Microcyclus	Slightly curved rods, also closed-ring stage	None	Aerobic, heterotrophic
Spirillum	S- to spiral-shaped	Tufts, polar	Aerobic, heterotrophic
Paraspirillum	S- to spiral-shaped	1, polar	Not cultured (aerobic ?, heterotrophic ?)
Selenomonas	Kidney- to crescent-shaped	Tufts, central	Anaerobic, heterotrophic
Myconostoc	Long, curved to comma-shaped	?, polar	Not cultured (aerobic, heterotrophic ?)

* As classified in *Bergey's Manual of Determinative Bacteriology*, 7th ed , Williams & Wilkins, 1957

acids provides energy for certain anaerobic, parasitic vibrios

A variety of simple organic compounds such as hexoses, mono- and dicarboxylic acids, and hydroxy and keto acids can be used as sole carbon and energy sources by many of the aerobic forms. More specialized abilities also exist, the genera *Cellvibrio* and *Cellfalcicula* are defined by their cellulose digesting ability, and various vibrios can utilize oxalate, agar, hydrocarbons, and aromatic compounds.

There are several human and animal pathogens in the genus *Vibrio*, including *V comma*, the causative agent of asiatic cholera (*see* CHOLERA VIBRIO), and *V. fetus*, the cause of abortion in cattle and sheep. Also associated with warm blooded animals is the *Selenomonas* group, found in the caecum, buccal cavity, or rumen These organisms have the vibrio form, but differ in having their flagella arising from the center of the concave side rather than polarly.

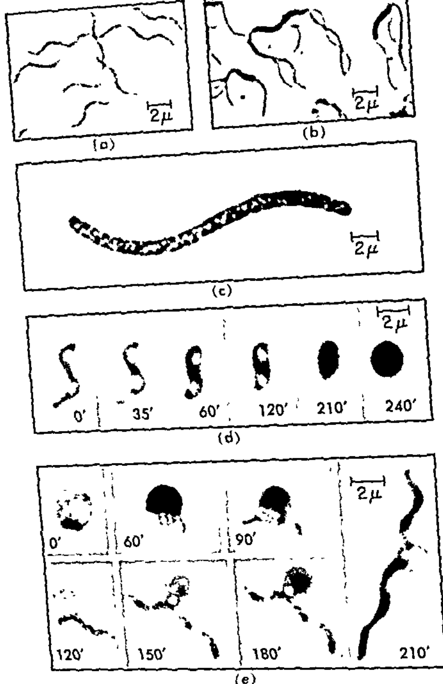
Desulvibrio desulfuricans. The most studied of the nonpathogenic *Spirillaceae* is *Desulfovibrio desulfuricans.* This organism grows either as a chemotrophic heterotroph or autotroph, oxidizing a variety of organic compounds or molecular hydrogen with the concomitant reduction of sulfate to hydrogen sulfide It is widely distributed in water, sediments, and soil, and is responsible for most of the sulfide in nature that is not of geothermal origin. The precipitation of iron sulfide by its activities results in the black sediments common in estuarine environments. Because of the strongly reducing environment resulting from sulfate reduction and other evidence, it is believed that *Desulfovibrio desulfuricans* plays a significant role in petroleum formation. In addition to its nuisance value as an odor producer, the organism also causes economic losses to man as an agent of anaerobic corrosion of piping, oil-well casings, and other buried structures. *See* PETROLEUM MICROBIOLOGY. [S.C.R.]

Bibliography: R S Breed et al, *Bergey's Manual of Determinative Bacteriology*, 7th ed., 1957; M. A. Williams, Some problems in the identification and classification of *Spirillum* I, Earlier taxonomy of the genus *Spirillum*, *Intern. Bull Bacteriol Nomenclature and Taxonomy*, 9:35–55, 1959

Spirochaetales

An order of bacteria characterized by elongate cells twisted three-dimensionally into a spiral shape Some members of the order are pathogens, causing syphilis (in man), relapsing fever (in man), Weil's disease (in man and other animals), while others are parasitic or free-living forms in water or sediments The pitch of the spiral varies so that different species may appear as tightly coiled, almost closed springs, as regular open coiled forms, or as irregularly twisted cells Morphologically similar forms exist among the larger spirilla, and intermediates between the two groups exist (see SPIRILLACEAE) The spirochetes are unusually flexible, being able to elongate and contract or to superimpose secondary waves on the primary coils of the spiral Flexibility, also shown by the slime bacteria, is attributed to a lack of the rigid cell wall typical of most bacteria (*see* MYXOBACTERALES).

All spirochetes are motile, usually swimming rapidly with a motion that involves both forward progress and spinning around the long axis Flagella are generally believed to be absent in this group, and the mechanism of motility remains es



(a) Typical spirillum (b) Flagellar stain showing spirillum with polar flagella (c) *Spirillum volutans* (d) Microcyst formation (e) Microcyst germination in *Spirillum lunatum* (time sequence of individual living cells, elapsed time in minutes) (Photographs by M A Williams; (d) and (e) from M. A. Williams and S. C. Rittenberg, Microcyst formation and germination in *Spirillum lunatum*, *J Gen. Microbiol*, 15 205, 1956)

sentially unexplained (see BACTERIAL MOTILITY). Reproduction is by the asexual process of transverse binary fission Some evidence also exists for the occurrence of life cycles in this group involving filterable forms and sexual stages, although most bacteriologists consider the evidence inconclusive. The order is divided into two families, Spirochaetaceae and Treponemataceae, primarily on the basis of size

Spirochaetaceae. This family includes the large spirochetes about 30–500 microns (μ) in length by 0 5–3 0 μ in cell diameter With the possible excep-

Genera of the family Spirillaceae*

Genus	Characteristic cell shape	Flagella, no. and position	Metabolism
Vibrio	Single curve, comma-shaped	1, polar	Aerobic to anaerobic, heterotrophic
Desulfovibrio	Single curve, comma-shaped	1, polar	Anaerobic, heterotrophic to autotrophic, reduces sulfate to sulfide
Methanobacterium	Straight to slightly curved	None	Anaerobic, heterotrophic to autotrophic, carbon dioxide reduced to methane
Cellvibrio	Slightly curved long rods	1, polar	Aerobic to facultatively anaerobic, heterotrophic, oxidizes cellulose
Cellfalcicula	Straight to curved, spindle-shaped	1, polar	Like *Cellvibrio*
Microcyclus	Slightly curved rods, also closed-ring stage	None	Aerobic, heterotrophic
Spirillum	S- to spiral-shaped	Tufts, polar	Aerobic, heterotrophic
Paraspirillum	S- to spiral-shaped	1, polar	Not cultured (aerobic ?, heterotrophic ?)
Selenomonas	Kidney- to crescent-shaped	Tufts, central	Anaerobic, heterotrophic
Myconostoc	Long, curved to comma-shaped	?, polar	Not cultured (aerobic, heterotrophic ?)

* As classified in *Bergey's Manual of Determinative Bacteriology*, 7th ed , Williams & Wilkins, 1957

acids provides energy for certain anaerobic, parasitic vibrios.

A variety of simple organic compounds such as hexoses, mono- and dicarboxylic acids, and hydroxy and keto acids can be used as sole carbon and energy sources by many of the aerobic forms More specialized abilities also exist; the genera *Cellvibrio* and *Cellfalcicula* are defined by their cellulose digesting ability, and various vibrios can utilize oxalate, agar, hydrocarbons, and aromatic compounds.

There are several human and animal pathogens in the genus *Vibrio*, including *V. comma*, the causative agent of asiatic cholera (*see* CHOLERA VIBRIO), and *V. fetus*, the cause of abortion in cattle and sheep. Also associated with warm blooded animals is the *Selenomonas* group, found in the caecum, buccal cavity, or rumen These organisms have the vibrio form, but differ in having their flagella arising from the center of the concave side rather than polarly.

Desulvibrio desulfuricans. The most studied of the nonpathogenic *Spirillaceae* is *Desulfovibrio desulfuricans* This organism grows either as a chemotrophic heterotroph or autotroph, oxidizing a variety of organic compounds or molecular hydrogen with the concomitant reduction of sulfate to hydrogen sulfide. It is widely distributed in water, sediments, and soil, and is responsible for most of the sulfide in nature that is not of geothermal origin. The precipitation of iron sulfide by its activities results in the black sediments common in estuarine environments Because of the strongly reducing environment resulting from sulfate reduction and other evidence, it is believed that *Desulfovibrio desulfuricans* plays a significant role in petroleum formation. In addition to its nuisance value as an odor producer, the organism also causes economic losses to man as an agent of anaerobic corrosion of piping, oil-well casings, and other buried structures *See* PETROLEUM MICROBIOLOGY. [S.C.R.]

Bibliography: R. S. Breed et al., *Bergey's Manual of Determinative Bacteriology*, 7th ed., 1957; M. A Williams, Some problems in the identification and classification of *Spirillum* I, Earlier taxonomy of the genus *Spirillum*, *Intern Bull Bacteriol Nomenclature and Taxonomy*, 9.35–55, 1959.

Spirochaetales

An order of bacteria characterized by elongate cells twisted three-dimensionally into a spiral shape Some members of the order are pathogens, causing syphilis (in man), relapsing fever (in man), Weil's disease (in man and other animals), while others are parasitic or free-living forms in water or sediments The pitch of the spiral varies so that different species may appear as tightly coiled, almost closed springs, as regular open-coiled forms, or as irregularly twisted cells Morphologically similar forms exist among the larger spirilla, and intermediates between the two groups exist (*see* SPIRILLACEAE) The spirochetes are unusually flexible, being able to elongate and contract or to superimpose secondary waves on the primary coils of the spiral Flexibility, also shown by the slime bacteria, is attributed to a lack of the rigid cell wall typical of most bacteria (*see* MYXOBACTERALES)

All spirochetes are motile, usually swimming rapidly with a motion that involves both forward progress and spinning around the long axis Flagella are generally believed to be absent in this group, and the mechanism of motility remains es

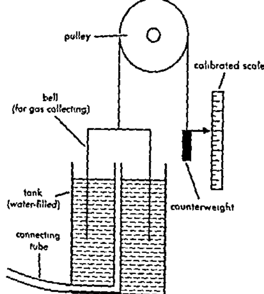


Fig 2 Subdivisions of the lung volume. (*From J R. Pappenheimer et al, Standardization of definitions and symbols in respiratory physiology, Federation Proc, 9(3) 602–605, 1950*)

not measurable by spirometry, is left in the lungs.

The volume of the lungs, when the muscles of breathing are completely relaxed, is the relaxation volume. At the end of ... balance ... this volume requires energy from natural (breathing muscles) or from artificial (mechanical respirator) sources See RESPIRATION, EXTERNAL.

[A B O.]

Spirotricha

A subclass of the class Ciliata which contains those ciliate Protozoa that are typified by conspicuous, compound ciliary structures. The structures are the cirri, which occur on the ventral surface of the body, and the buccal organelles, which include the undulating membrane and the adoral zone of membranelles This group of ciliates is classified into six orders Separate articles appear on each of these groups listed in the following classification:

Subclass Spirotricha
  Order Heterotrichida
    Odontostomatida
    Oligotrichida
    Tintinnida
    Entodiniomorphida
    Hypotrichida

The general body ciliation is much reduced, or even entirely absent, except in one order The body size is usually greater than that of species belonging to the subclass Holotricha The orders in this subclass contain those organisms which are cons... the most ...

Spiruroidea

An order of the class Nematoda. These roundworms are parasites of the alimentary canal, respiratory system, or orbital, nasal, and oral cavities of vertebrates All of man's domestic animals are subject to parasitism by one or more members of this group. A few species are accidental parasites of man. In the alimentary tract they cause gastroenteritis, peritonitis, and more or less serious hemorrhage and inflammation. Some species cause tumors and nodules, others infect the eyes of their hosts, sometimes causing blindness.

General morphology. These worms may be slender and threadlike, large and heavy bodied, or short and thick. The sexes are dimorphic in some species. The mouth is usually surrounded by a single pair of lateral pseudolabia, and interlabia are sometimes present The buccal cavity may be well developed or rudimentary. The esophagus is divisible into two parts, a shorter anterior muscular portion and a longer posterior glandular portion which connects with the simple intestine. Spines or other cuticular adornments are sometimes present on the head and body The male has two dissimilar spicules, unequal in length, and a tail usually with broad winglike processes (alae) which are often ornamented with cuticular markings and provided with stalked papillae. The vulva of the female usually opens near the middle of the body, sometimes posteriorly, oc...

...lowing the intermediate host containing the infective stage of the parasite.

Important spiruroids. Usually the spiruroids are most important as parasites of domestic animals, although human infestations have been recorded. The following table includes some of the more common spiruroids of domestic animals

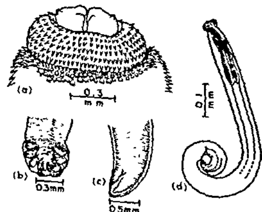


Fig 1 *Gnathostoma spinigerum* (a) Head (b) Tail of male. (c) Tail of female. (d) Entire animal (*from H. Baylis and C Lane, 1920*). (*Parasito...*



1 Spirochaeta (Spirochaetaceae) 4 Borrelia (Treponemataceae)
2 Saprospira (Spirochaetaceae) 5 Treponema (Treponemataceae)
3 Cristispira (Spirochaetaceae) 6 Leptospira (Treponemataceae)

Representative genera of the Spirochaetales. (W. E. D. Stewart)

tion of *Spirochaeta plicatilis*, these forms have not been grown in pure culture and little is known of their physiology and metabolism. The family consists of three genera, *Spirochaeta*, *Cristispira*, and *Saprospira*, that are distinguished both on morphological and ecological bases.

Spirochaeta species occur in both fresh and marine waters, usually in the presence of decaying organic matter. Since no specific technique has been worked out for enriching the medium so as to procure their growth selectively in a mixed culture, their detection is somewhat a matter of chance. They were originally described as possessing an axial filament, not digested by trypsin, around which the protoplast is spirally wound. It has recently been claimed, however, that the axial filament is an artifact of the staining procedure.

Cristispira species occupy a peculiar ecological niche, being found only within the crystalline style of certain mollusks. They have a thin membrane, called the crista, fused to the cell surface which undulates out of phase to the spiral of the cell proper. The function of the crista is unknown although it has been postulated to play a role in motility.

Saprospira species resemble *Cristispira* species in general form but lack the crista. The first are found in marine sediments and the second in the intestinal tracts of mollusks.

Treponemataceae. This family includes the spirochetes less than 20 μ long and under 0.5 μ in diameter. Most described species are parasitic, including some important human pathogens. Free-living forms are commonly observed in water and infusions, but these have not received much study. The parasitic and pathogenic species that have been cultured have complex growth requirements which are usually satisfied by the addition of animal tissues and body fluids to culture media. Some have recently been grown in synthetic media. Both aerobic and anaerobic species occur. Some forms, in particular the syphilis organism, fail to stain with aniline dyes, and many are quite sensitive to organic arsenical compounds. These properties are not shared by most bacteria.

The family is divided into three genera that can be distinguished morphologically.

Borrelia species have coarse, irregular spirals and cause relapsing fever (*B. recurrentis*) and Vincent's angina (*B. vincentii*). See RELAPSING FEVER; VINCENT'S ANGINA.

Treponema species have uniform, somewhat angular spirals and cause syphilis (*T. pallidum*), yaws (*T. pertenue*), and pinta (*T. carateum*). See PINTA; SYPHILIS; YAWS.

Leptospira species have tightly coiled, almost closed spirals; *L. icterohaemorrhagiae* causes Weil's disease. See BACTERIA, TAXONOMY OF; SCHIZOMYCETES. [S.C.R.]

Spirochete

A member of a group of microorganisms distinguished by their spiral form and active motility. There are three groups of medical importance, based on morphological criteria: *Treponema*, the cause of syphilis and related treponematoses; *Borrelia*, the cause of relapsing fever and implicated in Vincent's angina; *Leptospira*, the cause of one form of infectious hepatitis, known as Weil's disease, as well as other leptospiroses. Each group contains nonpathogenic varieties which are indistinguishable morphologically from those producing human and animal diseases.

In addition, several species of large nonpathogenic spirochetes (500 μ in length) are found in shellfish, stagnant water, and other wet environments. See BACTERIOLOGY, MEDICAL; BEJEL; LEPTOSPIRA; RAT-BITE FEVER; RELAPSING FEVER; SPIROCHAETALES; SYPHILIS; VINCENT'S ANGINA; WEIL'S DISEASE; YAWS. [T.B.T.]

Bibliography: R. Dubos (ed.), *Bacterial and Mycotic Infections of Man*, 3d ed., 1958; T. B. Turner and D. H. Hollander, *Biology of the Treponematoses*, World Health Organization Monograph 35, 1957.

Spirometry

The measurement, by a form of gas meter, of volumes of air that can be moved in or out of the lungs (Fig. 1). The volume that moves in and out with each breath is the tidal volume. The maximal possible tidal volume is the vital capacity. Other subdivisions of lung volume are indicated in Fig. 2. Even after the most complete possible expiration, a considerable volume of gas, the residual volume,

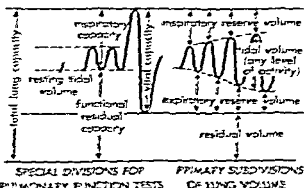


Fig. 1. Diagram of a spirometer.

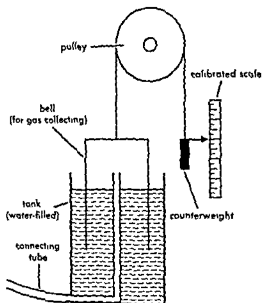


Fig 2 Subdivisions of the lung volume (*From J R. Pappenheimer et al, Standardization of definitions and symbols in respiratory physiology, Federation Proc, 9(3) 602–605, 1950*)

not measurable by spirometry, is left in the lungs

The volume of the lungs, when the muscles of breathing are completely relaxed, is the relaxation volume. At this volume, approximately that at the end of normal expiration, elastic forces of the lung balance those of the chest wall Displacement from this volume requires energy from natural (breathing muscles) or from artificial (mechanical respirator) sources. *See* RESPIRATION, EXTERNAL.

[A B O]

Spirotricha

A subclass of the class Ciliata which contains those ciliate Protozoa that are typified by conspicuous, compound ciliary structures The structures are the cirri, which occur on the ventral surface of the body and the buccal organelles, which include the undulating membrane and the adoral zone of membranelles This group of ciliates is classified into six orders Separate articles appear on each of these groups listed in the following classification:

Subclass Spirotricha
 Order Heterotrichida
  Odontostomatida
  Oligotrichida
  Tintinnida
  Entodiniomorphida
  Hypotrichida

The general body ciliation is much reduced, or even entirely absent, except in one order The body size is usually greater than ... the subcla... contain the ... the most ...

Spiruroidea

An order of the class Nematoda These roundworms are parasites of the alimentary canal, respiratory system, or orbital, nasal, and oral cavities of vertebrates All of man's domestic animals are subject to parasitism by one or more members of this group. A few species are accidental parasites of man. In the alimentary tract they cause gastroenteritis, peritonitis, and more or less serious hemorrhage and inflammation. Some species cause tumors and nodules, others infect the eyes of their hosts, sometimes causing blindness.

General morphology. These worms may be slender and threadlike, large and heavy-bodied, or short and thick The sexes are dimorphic in some species. The mouth is usually surrounded by a single pair of lateral pseudolabia, and interlabia are sometimes present. The buccal cavity may be well developed or rudimentary. The esophagus is divisible into two parts, a shorter anterior muscular portion and a longer posterior glandular portion which connects with the simple intestine. Spines or other cuticular adornments are sometimes present on the head and body. The male has two dissimilar spicules, unequal in length, and a tail usually with broad winglike processes (alae) which are often ornamented with cuticular markings and provided with stalked papillae. The vulva of the female usually opens near the middle of the body, sometimes posteriorly, occasionally anteriorly.

Life cycle. The life cycle of all spiruroids thus far studied is indirect, that is, an intermediate host is involved which is usually some kind of an arthropod. The definitive host becomes infected by swallowing the intermediate host containing the infective stage of the parasite.

Important spiruroids. Usually the spiruroids are most important as parasites of domestic animals, although human infestations have been recorded The following table includes some of the more common spiruroids of domestic animals.

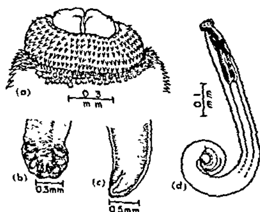


Fig 1 *Gnathostoma spinigerum. (a) Head (b) Tail of male (c) Tail of female. (d) Entire animal (From H Baylis and C. Lane, 1920) (Parasite Laboratory, Agricultural Research Service, Beltsville, Maryland)*

Gnathostoma spinigerum. This is a robust worm, 25–50 mm long, with the head end developed into a globular swelling armed with eight or more rows of hooks. Behind the head, the body is covered for about half its length with overlapping rows of toothed scales (Fig. 1). Its normal hosts among domestic animals are dogs and cats, in which it produces and inhabits large tumors in the stomach wall. These worms may cause a fatal peritonitis. The primary larval host is the fresh-water crustacean, *Cyclops*, and the second intermediate host is a fish, amphibian, or snake. When one of these vertebrate animals swallows an infected *Cyclops*, the *Gnathostoma* larva invades the tissues of its new host and becomes encysted. The second intermediate host is obligatory, for carnivorous mammals do not become infected from ingesting infected *Cyclops*. Human gnathostome infections are of frequent occurrence in southeast Asia. They result from eating raw or improperly cooked fish, especially eels and frogs, all of which are commonly infected. When ingested by man, the larvae develop to adult but sexually immature worms. Instead of invading the stomach wall, they wander around in

Common spiruroid parasites of domestic animals

Parasite	Definitive host	Intermediate host	Distribution	Pathology
Ascarops strongylina (Rudolphi, 1819)	Swine	Beetles	Cosmopolitan	Inflammation of stomach mucosa
Cheilospirura hamulosa (Diesing, 1851)*	Fowl, turkey	Grasshoppers, beetles	N. America, S. America, Asia, Europe	Soft nodules in musculature of gizzard
Dispharynx nasuta (Rudolphi, 1819)	Fowl, turkey, pigeon, guinea fowl, pheasant	Isopods	N. America, S. America, Europe, Asia	Ulcers and gland destruction in proventriculus
Echinuria uncinata (Rudolphi, 1819)	Duck, goose, swan	Water fleas	N America, Europe, Asia, N. Africa	Inflammation and nodules in proventriculus and gizzard
Gnathostoma hispidum Fedtchenko, 1872	Swine		Europe, Asia, S Africa, Australia	Destruction of liver tissue, gastric lesions
Gnathostoma spinigerum Owen, 1836*	Cat, dog	*Cyclops;* Fish and frogs	Asia, Australia, N America	Destruction of liver tissue, gastric lesions
Gongylonema pulchrum Melin, 1857*	Horse, cattle, sheep, goat, swine, camel, zebu, buffalo	Beetles	Cosmopolitan	In esophagus (unimportant)
Habronema megastoma (Rudolphi, 1819)	Equines	*Musca* spp	Europe, Asia, Africa, Australia, S America	Tumors in stomach wall
Habronema microstoma (Schneider, 1866)	Equines	Stable fly, house fly	Cosmopolitan	Catarrhal gastritis, stomach ulcers
Habronema muscae Carter, 1861	Equines	*Musca* spp	Cosmopolitan	Catarrhal gastritis
Oxyspirura mansoni (Cobbold, 1879)	Fowl, turkey, peafowl	Cockroaches	Southern U S, S. Asia, Australia, S America	Lesions in eyes, blindness, loss of eyeball
Physaloptera spp.	Dog, cat	Unknown	Cosmopolitan	Erosion and inflammation of stomach mucosa
Physocephalus sexalatus (Molin, 1860)	Swine, camel, rabbit	Beetles	Cosmopolitan	Inflammation of stomach mucosa
Spirocerca lupi (Rudolphi, 1809)	Dog	Beetles	Cosmopolitan	Hemorrhage and abscesses in stomach wall, aorta, esophagus
Tetrameres americana Cram, 1927	Fowl, turkey	Grasshoppers, cockroaches	N America, S America, Africa, Asia, Europe	Irritation and inflammation of proventriculus
Thelazia californiensis Price, 1930	Sheep, cat, dog	Unknown	California	Scar tissue formation in eyes, blindness
Thelazia callipaeda Railliet and Henry	Dog, rabbit	Unknown	Far East	Scar tissue formation in eyes blindness
Thelazia rhodesii (Desmarest, 1828)	Cattle, sheep, goat, buffalo	*Musca* spp	Europe, Asia, Africa	Scar tissue formation in eyes, blindness

* Occasional parasite of man

and under the skin, and are the cause of migrating intermittent swellings or *edema*, and sometimes a creeping eruption. Eventually they become encapsuled or escape through an abscess

Thelazia The species of *Thelazia* (Fig 2) are small, slender worms, less than 20 mm in length that possess a cuticle with prominent cross striations In the female the vulva is anterior and the male is without caudal alae. The worms inhabit the conjunctival sac and lachrymal ducts of the eye. At times, they creep out over the eyeball and later return to the inner corner of the eye. They irritate the eye by their movements and, in some cases, cause blindness The intermediate hosts, where known, are flies of the genus *Musca*, which cluster around the eyes of infected animals and doubtless

is smooth and in the female the vulva is located posteriorly (Fig 3) Cockroaches serve as intermediate hosts of *O mansoni*, the eyeworm of poultry. When an infected cockroach is eaten by a fowl, the nematode larva wanders up the esophagus,

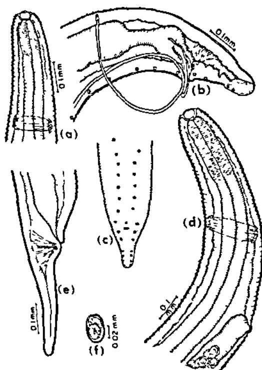


Fig 2 *Thelazia platyptera* (a) Anterior end of male, lateral view (b) Caudal end of male, lateral view (c) Caudal end of male, ventral view (diagrammatic) (d) Anterior end of female, lateral view. (e) Caudal end of female, lateral view (f) Egg (from J. Hwang and E Wehr, 1957) (Parasite Laboratory, Agricultural Research Service, Beltsville, Maryland)

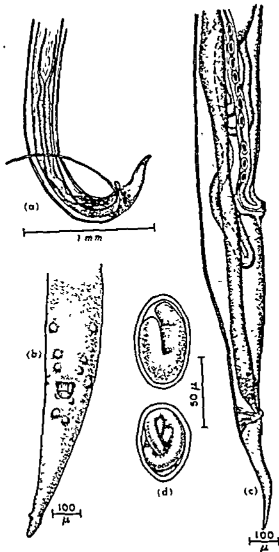


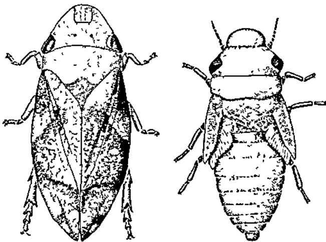
Fig. 3 *Oxyspirura mansoni.* (a) Male tail, lateral view (b) Male tail, ventral view. (c) Female tail. (d) Eggs (after B Ransom, 1904) (Parasite Laboratory, Agricultural Research Service, Beltsville, Maryland)

pharynx, and lachrymal duct to the eye where it may cause blindness and in severe cases complete destruction of the eyeball The life cycle of the nematode is completed when its eggs are washed down the tear ducts, swallowed, and passed out in the droppings

Other spiruroids Among the many genera included in this group that are on occasion of importance as parasites of domestic animals are *Spirocerca*, *Habronema*, *Physocephalus*, *Gongylonema*, *Streptocara*, and *Tetrameres*. [J L C]

Spittle insect

Any member of the family Cercopidae, order Homoptera. They are also called spittlebugs or froghopper, the latter because of the slight resemblance of some species to miniature frogs. These insects are not more than 1/4 in long, dark in color, usually brown. They derive their name from the white, frothy mass which the female deposits on



The spittle insect, *Lepyronia quadrangularis*; length to ⅓ in. (*From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

plant stems with each egg. The nymph develops in this frothy mass, which provides a moist environment and concealment from possible enemies There is only one generation a year. When abundant, spittle insects can cause serious stunting of plants, especially clover. *See* HOMOPTERA; INSECTA. [J.D.B.]

Spleen

An organ present in most vertebrates which lies in the abdominal cavity, usually in close proximity to the left border of the stomach. It is subject to wide variation in size, shape, color, and location, depending upon the species and the age of the individual.

Normally, a human spleen measures about 1 by 3 by 5 in and weighs less than ½ lb. It is a firm organ with an oval shape and is indented on its inner surface to form the hilum, or stalk of attachment to the peritoneum These folds of mesentery also carry the splenic artery and vein to the organ.

The spleen is composed of a bloody pulp lying between fibrous partitions, or trabeculae These partitions originate from the dense fibrous capsule that surrounds the organ

The circulation is unusual in that the splenic arterioles open into thin-walled dilations, called sinusoids, which in turn drain into small veins This arrangement of distensible blood-filled vessels, clusters of cells, and developing follicles accounts for the pulpy character of the spleen

In this pulp are many kinds of both red and white blood cells Microscopically, the major feature of the pulp is the presence of Malpighian corpuscles, or follicles, which consist of aggregations of developing lymphocytes These follicles form initially around the small branches of the splenic artery. In addition to lymphocytes, other white cells such as monocytes, histiocytes, and giant cells are found in quantity. Red blood cells are freely mixed with the white cells, except in the developing follicles, where they are usually crowded out *See* BLOOD.

The spleen is an important part of the blood-forming or hematopoietic system; it is also one of the largest lymphoid organs in the body, and as such is involved in the defenses against disease attributed to the reticulo-endothelial system. This system consists of many different types of cells in the body which have the power to neutralize or engulf foreign particles or bacteria, and sometimes act as filters through which blood or lymph must pass.

Although the chief functions of the spleen appear to be the production of lymphocytes, the probable formation of antibodies, and the destruction of "worn out" red blood cells, other less well understood activities are known. For example, in some animals, it may act as a blood reservoir and is said to contract rhythmically at frequent intervals, thus aiding in the return of blood through the liver to the heart. In the fetus, and sometimes in later life, the spleen may also be a primary center for the formation of red blood cells. *See* ANTIBODY.

Another function of the spleen is its role in biligenesis. Because the spleen destroys erythrocytes, it is one of the sites where extrahepatic bilirubin is formed. Bilirubin is the principal pigment of bile, formed by the reduction of biliverdin. Bile is normally present in the feces.

In lower vertebrates, splenic tissue appears as scattered lymphoid masses in the wall of the digestive tract and apparently acts as the primary blood-forming organ. Bone marrow appears in certain amphibians and in most higher forms, taking over the formation of red cells in adult forms under normal conditions. The spleen then produces cells which are principally of the lymphoid type, and serves in other ways, some of which have been mentioned. This organ is not necessary to life because its activities may be shared or assumed by other organs under proper conditions. Many of the splenic functions, such as lymphocyte formation, are common to all lymphoid tissues *See* SPLEEN DISORDERS. [E.G.ST.]

Spleen disorders

The spleen is rarely the site of primary disorders except those of vascular origin, but it is frequently involved in systemic inflammations, metabolic diseases, and generalized blood disorders.

Among vascular disturbances, acute and chronic congestion are prominent, particularly chronic congestion caused by cardiac failure, cirrhosis of the liver, and obstruction of the blood flow from the spleen by thrombi, scarring, or tumor tissue Obstruction of the splenic artery or its branches by thrombi may result in an infarct caused by either cardiac or blood disease.

Inflammations include acute and chronic forms The characteristic engorgement of blood often causes a marked enlargement of the organ Bacteremias frequently produce this enlargement, or splenomegaly, and inflammation, but any severe infectious disease such as diphtheria or pneumonia may do so Tuberculosis, syphilis, typhoid fever, and malaria, as well as many other specific infections may cause splenomegaly, often with char-

acteristic gross or microscopic changes for each disease. *See* MALARIA; SYPHILIS; TUBERCULOSIS; TYPHOID FEVER

The spleen is involved in certain types of pigment metabolism and is frequently the site of specific kinds of degeneration when amyloidosis, hemosiderosis, or argyrosis is present.

Similarly, in disorders of lipid metabolism, abnormal kinds or amounts of fat may appear in connection with Gaucher's disease, the hereditary Neimann Pick disease, and others. *See* LIPID METABOLISM.

Because the spleen normally aids in the destruction of worn-out red blood cells and, under certain conditions, in the formation of red cells, it is not uncommon to find splenic involvement in blood disorders such as sickle-cell anemia, congenital hemolytic jaundice, polycythemia, and Mediterranean anemia *See* HEMATOLOGIC DISORDERS

The leukemias, especially when of the lymphocytic or neutrophilic varieties, cause some of the most prominent cases of splenomegaly as well as other changes In myeloid leukemia, for instance, spleen weight of 6000–8000 grams is not rare and the spleen may fill the entire abdomen.

Hypersplenism is an obscure disorder in which one or more of the blood cell types are destroyed to excess Primary hypersplenism results from unknown causes, secondary hypersplenism may follow inflammatory diseases, chronic congestion, or tumor invasion

Tumors originating in the spleen are rare and usually limited to such benign growths as hemangiomas, lymphangiomas, and fibromas, but malignant lymphomas and lymphosarcomas also occur

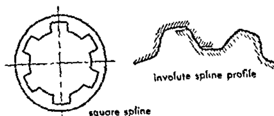
Secondary tumors, which originate elsewhere and metastasize to the spleen, are not uncommon, particularly when of the lymphoma group Other varieties are seen less often and include carcinomas, particularly either from a blood stream invasion or from growth in the neighboring stomach or intestines

Trauma to the spleen is more common than suspected and the consequences of rupture of the organ are often tragic because this may follow an apparently moderate abdominal injury.

Many other conditions produce splenic change, notably splenomegaly of some degree The presence of an enlarged spleen indicates that a thorough diagnostic evaluation should be made. *See* SPLEEN [E.C.ST]

Splines

A series of projections and slots used instead of a key to prevent relative rotation of cylindrically fitted machine parts Splines are several projections machined on the shaft, the shaft fits into a mating bore called a spline fitting Splines are made in two forms square and involute (as illustrated). Since there are several projections (integral keys) to share the force in transmitting power, the splines can be shallow, thereby not weakening the shaft as much as would a standard key



Square spline, and profile of involute spline.

Square splines have 4, 6, 10, or 16 splines. The external part (shaft) may have the splines formed by milling and the internal part (bore) by broaching Three classes of fits are used: sliding (as for gear shifting) under load, sliding when not loaded, and permanent fit. Square splines have been used extensively for machine parts. In the automotive industry, square splines have been replaced generally by involute splines, which cost less to make accurately for good fit and interchangeability

Involute splines are used to prevent relative rotation of cylindrically fitted machine parts and have the same functional characteristics as square splines. The involute spline, however, is like an involute gear, and the spline fitting (internal part) is like a mating internal gear. Profiles are the same as for gear teeth of the stub (fractional pitch) form with 30° pressure angle.

Involute splines on the shaft are generated by a hob or a gear shaper, and internal splines are formed by a broach or a gear shaper. Three classes of fit are standard, sliding, close, and press.

Involute serrations are similar to involute splines except that the pressure angle is 45° and, while there are three standard fits (loose, close, and press), serrations are usually press fitted and used for permanent assembly. They are used for both parallel and tapered shafts. *See* MACHINE KEY. [P.H.B]

Bibliography. American Standards Association, *Involute Splines, Side Bearing*, B5 15–1950, ASA, *Involute Serrations*, B5.26–1950; ASA, *Involute Spline and Serration Gages and Gaging*, B5.31–1953

Spodumene

The name given to the monoclinic lithium pyroxene $LiAl(SiO_3)_2$. Spodumene commonly occurs as white to yellowish prismatic crystals, often with a "woody" appearance, exhibiting the 87° pyroxene (110) cleavages. It is easily identified during heating in a flame by the red color given off, accompanied b

m

h

m

w accompanied by a 30% increase in volume Spodumene is capable of forming immense crystals in nature. A single crystal 47 ft in length and 5 ft in diameter, and others almost as large, have been found at the Etta mine in South Dakota. This implies the remarkable ability of a single crystal to replace a large variety of preexisting minerals and

yet maintain the integrity of a single crystal, a crystal growth that is unequaled elsewhere in nature. *See* PYROXENE.

Spodumene is usually found as a constituent in certain granitic pegmatites in association with quartz, alkali feldspars, mica, beryl, phosphates and a large variety of rare minerals. It is also known to occur as disseminated grains in some granite gneisses. Spodumene often alters to a fibrous mass composed of eucryptite $LiAlSiO_4$ and albite, or eucryptite and muscovite. The emerald-green variety, hiddenite, and a lilac variety, kunzite, are used as precious stones. Spodumene from pegmatites is used as an ore for lithium. *See* LITHIUM. [C.W.D.]

Sponge

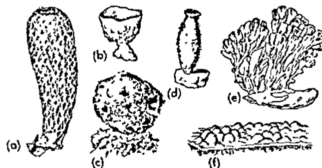
Any of about 5000 species comprising the phylum Porifera. Except for a fresh-water family of 150 species, all are marine. Three classes are recognized, based on skeletal type. The skeletons of the class Calcarea have 1-, 3-, or 4-branched spicules of calcium carbonate Most of these live in shallow water, are less than 6 in long, and are relatively simple. The glass sponges, class Hexactinellida, have 6-rayed spicules of siliceous material, arranged either separately or in bunches. The bath sponges, class Demospongiae, typically have some fibrous material, called spongin, in their skeletons. However, they may lack either spongin, spicules, or both. The fresh-water sponges also belong to this class.

Economic importance. Although now largely replaced by synthetic materials, sponges are still of some commercial importance. Most bath sponges come from the Gulf of Mexico, especially that part bordering Florida and the West Indies. Sponge farming is practiced in the Bahamas by cutting up large sponges and planting the small pieces in favorable habitats. There is also a substantial commercial production of sponges in the Mediterranean Sea, where the Greeks have harvested them for centuries

Biological significance. In addition to their commercial use, sponges are of interest to biologists because of their simple structure, organization, and regenerative powers. Sponges also serve as homes for large numbers of small marine animals. A very large sponge may shelter more than 16,000 animals, mostly small shrimp and crabs, but also a wide variety of other invertebrates and fishes

Size and structure. Sponges vary in size from 1 mm to 6 ft in diameter. They are mostly asymmetrical, but a few, for example the large vaselike forms, show radial symmetry. Some are flat, many are globular, and others are branched. Most sponges are drab in color, but some are yellow, red, blue, or black. In the sea they are found from low tide to a depth of 3½ miles.

Sponges are the simplest of the many-celled animals, closely resembling some of the colonial, flagellate Protozoa. Their organization is that of a loose aggregate of cells without definite germ



Types of sponge. (a) *Regadrella*, glass sponge, class Hexactinellida (after Lankester), (b) *Poterion*, Neptune's goblet, class Demospongiae, (c) *Euspongia*, bath sponge, class Demospongiae; (d) *Scypha* (formerly called *Sycon*), class Calcarea (after Lankester); (e) *Microciona*, class Demospongiae; (f) *Haliclona*, encrusting sponge, class Demospongiae (*From T. I. Storer and R. L. Usinger, General Zoology, 3d ed, McGraw-Hill, New York, 1957*)

layers. A simple sponge has an outer layer of somewhat generalized cells supporting a layer of collar cells, each bearing a flagellum. Pores through the body wall permit the passage of water into the inner cavity, or spongocoel, the current being created by the beating of the flagella. Water exits by means of a large terminal opening, the osculum. More complex sponges show elaborate modifications of the water circulation system. Plankton trapped by the flagella are either digested by the collar cells or passed to a nearby cell.

Reproduction. Reproduction occurs sexually, by budding, by gemmule formation (a type of internal budding), or by regeneration. Sponge cells have remarkable regenerative capacity. Sponge cells regroup to form new animals even when rubbed through screening silk. Most sponges are hermaphroditic, but male and female cells are seldom produced at the same time. Generalized body-wall cells become modified into gametes. The egg remains in the body wall where it begins development after fertilization. The embryo develops into a cluster of cells, some of which have flagella After its escape from the parent, the embryo is free-swimming for a period, then inverts, the flagellated collar cells thus being turned inside Soon it attaches itself to a solid substrate and matures into an adult. *See* PORIFERA; REGENERATION (BIOLOGY) [J.D.B.]

Sporobolomycetales

An order of fungi whose taxonomic position has long been in doubt. There is now evidence that they belong to the class Basidiomycetes The single family Sporobolomycetaceae contains six genera, *Sporobolomyces, Bullera, Tilletiopsis, Itersonilia, Sporidiobolus*, and *Dacromyces*. Only the first two genera are clearly yeastlike; the others are moldlike The family is characterized by the formation of mycelium, pseudomycelium, and budding yeast cells. Some of the vegetative cells form aerial sterigmata upon which the sexual ballistospores (a type of basidiospores) are formed. When mature,

the spores are discharged with great force by a drop-excretion mechanism. In the genus *Sporobolomyces* the ballistospores are asymmetrical (sickle- or kidney-shaped), whereas in *Bullera* they are symmetrical (oval or spherical). In addition, the species of *Sporobolomyces* are pink or salmon-colored, the *Bullera* cream-colored to pale yellow. They may be considered as the perfect stages of corresponding *Rhodotorula* species. Species of both genera have budding cells and all species are strictly oxidative. See BASIDIOMYCETES; YEAST.

[H.J.P.]

Bibliography: J. Lodder and N. J. W. Kreger van Rij, *The Yeasts—a Taxonomic Study*, 1952.

Sporotrichosis

A mycotic infection of man caused by *Sporotrichum schenckii*, a fungus that is world-wide in distribution. This organism has been isolated from timber, plants, and soil. The disease is most frequently seen in farmers and horticulturists who usually acquire the infection by injuring a finger or hand with contaminated plant materials. The primary lesion develops within 2-8 weeks following introduction of the fungus into the tissue, and appears as a hard nodule which is pink in color. Gradually, the lesion becomes darker in color and undergoes necrosis with discharge of purulent material. This lesion is referred to as a sporotrichotic chancre. As the infection progresses, numerous nodules are seen to develop along the lymphatic chain of the arm. Rarely does the disease progress beyond this form. Although the primary lesion may heal, the secondary nodules may persist for months to years if untreated.

Sporotrichum schenckii is a diphasic fungus, being a mold in nature, but becoming a *budding* yeast when growing in tissue. Although it is very difficult to demonstrate the organism in histological preparations of pus or tissue, it is readily cultured on most laboratory media. When inoculated in glucose cystine blood agar and incubated at 37°C, the yeast phase is seen. When cultures are incubated at room temperature the mold phase, which is buff to black in color and leathery in consistency, is obtained.

Potassium iodide given orally is the drug of choice, and it should be continued for several weeks following clinical recovery. See MYCOLOGY, MEDICAL.

[L.D.H.]

Sporozoa

A subphylum of the phylum Protozoa. These animals lead an entirely parasitic existence. The lack of organs of locomotion such as pseudopods, cilia, or flagella distinguishes this group from the other subphyla. No structure is present for the capture or ingestion of food particles. Nutrients are obtained in liquid form by endosmosis. Propagation is by the formation of more or less hard-shelled resistant spores, inside of which either sickle-shaped sporozoites or an ameboid sporoplasm develops. After escape from the spores, these infective stages become parasitic in a new host.

Originally, the importance of the spores was emphasized, however, it was realized that the spores of gregarines and coccidians are not homologous with those of the Myxosporidia. Fundamental differences in the life cycles were also known. With the elucidation of the general life histories of the malarial and related parasites, the role of the spore was deemphasized. It was noted that with the interjection of an intermediate host in the life history of the parasite, the necessity for spores was eliminated. Emphasis was shifted to the sporozoites as the infective stage, irrespective of how they were produced. Thus F. Schaudinn proposed the subclass Telosporidia to include the Haemosporidiida, Gregarinida, and Coccidia and the subclass Neosporidia to comprise the Myxosporidia, Microsporidia, and Actinomyxidia, whose spores bear polar capsules. For these reasons, it is difficult to characterize the subphylum Sporozoa. Still to be mentioned are such heterogeneous groups such as the Helicosporida, *Babesia*, *Theileria*, *Toxoplasma*, *Dactylosoma*, and other [illegible]

[illegible] of diverse origins whose only bond is that all are parasitic and produce spores or are related to forms that do.

Taxonomy. The Sporozoa is divided into three classes, the Telosporidea, Cnidosporidea, and Acnidosporidea. The Telosporidea are the most typical Sporozoa because their life cycles offer a good balance between schizogony and sporogony. Their spores lack polar capsules, and the sporozoites are formed inside them except in the Haemosporidiida. Cnidosporidea have spores with polar capsules; however, if fertilization occurs it is not followed by sporogony [illegible]

mc [illegible]

lu [illegible]

ab [illegible] of the liquefied protoplasm of the host, or of its food in solution, through their cell walls or membranes. The trophozoites of the microsporidian *Nosema* [illegible]

cles of [illegible]

polymo [illegible]

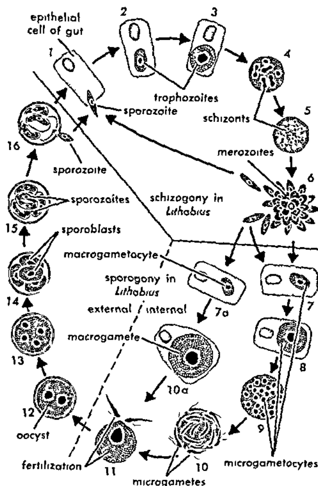
has re [illegible]

parasit [illegible]

of the [illegible]

brane [illegible] them in the food vacuoles so formed, a process for which the name phagotrophy was proposed. It is possible that further study will reveal that Sporozoa are capable of ingestion of particulate food to a greater extent than formerly supposed. The oocysts belonging to the coccidium *Eimeria wierzejskii* of the carp's intestine, found within the cytoplasm of Myxosporidia presumably after phagocytosis, are probably not digested.

Life cycles. The life of a sporozoan is basically a cycle, or a stereotyped series of stages and processes which are repeated in the same order in the next cycle. Its significance is the perpetuation of the species through reproduction and dissemination of infective stages.



Life history of Eimeria schubergi parasitic in the intestine of the centipede Lithobius. The centipede ingests oocysts from which the sporozoites are liberated in the intestine where they invade epithelial cells and develop into trophozoites. Trophozoites undergo schizogony with merozoites resulting. Schizogony may be repeated several times. Some merozoites eventually transform into gametocytes with fertilization and zygote formation resulting. The zygote produces the oocyst in which sporoblasts and sporozoites are formed. Oocysts develop outside the host's body and are ingested by another centipede. (From T. L. Jahn and F. F. Jahn, How to Know the Protozoa, Brown, 1949)

The basic life cycle of a typical species of Telosporidea consists of two phases: the asexual, in which there is multiplication of presumably nonsexual forms by schizogony, and, following schizogony, multiplication of patently sexual forms by sporogony.

The diversity of life cycles among the Cnidosporidea makes extensive generalization almost impossible, but it is safe to say that sporogony does not occur in this group, and that the infective stage which leaves the spore and becomes a trophozoite in the new host is an ameboid sporoplasm, not an elongate sporozoite capable of flexion and gliding movements. So far as the class Acnidosporidea is concerned, too little is known regarding the affinities and life cycles in its subclass Sarcosporidia, and knowledge concerning the life cycles for species of the subclass Haplosporidia is meager, though it is known that true sporogony does not occur and that in certain species an ameboid organism escapes from the spore after it is ingested by a new host.

Distribution. The hosts of Sporozoa include representatives of practically every branch of the animal kingdom. Gregarines are known to occur in coelenterates, echinoderms, flat worms, annelids, arthropods, mollusks, and certain lower chordates. Hyperparasitism is not unknown. For example, the microsporidian *Nosema helminthorum* is parasitic in the tapeworms *Moniezia expansa* and *Moniezia benedeni* which infest the intestines of sheep. Several other species of the same genus are known to parasitize other parasitic protozoa such as ciliates, cephaline or acephaline gregarines, and myxosporidians. Since the host range of a particular sporozoan species is extremely limited, Sporozoa are said to exhibit a high degree of host-specificity. *See* ACNIDOSPORIDEA; CNIDOSPORIDEA; PROTOZOA; TELOSPORIDEA. [E R BE.]

Spot welding

A resistance-welding process in which coalescence is produced by the flow of electric current through the resistance of metals held together under pressure. A low-voltage, high-current energy source is required (see Fig. 1). Usually the upper electrode moves and applies the clamping force. Pressure must be maintained at all times during the heating cycle to prevent flashing at the electrode faces. Electrodes are water-cooled and are made of copper alloys, because pure copper is soft and deforms under pressure. *See* RESISTANCE WELDING.

The electric current flows through at least seven resistances connected in series for any one weld. They are (1) upper electrode, (2) contact between upper electrode and upper sheet, (3) body of upper sheet, (4) contact between interfaces of sheets, (5) body of lower sheet, (6) contact between lower sheet and electrode, and (7) lower electrode (see Fig. 2). Heat generated in each of the seven sections will be in proportion to the resistance of each. The greatest resistance is at the interfaces (4) and heat is most rapidly developed there. The liquid-cooled electrodes 1 and 7 rapidly dissipate the heat generated at the contact between electrodes and sheets 2 and 6 and thus contain the

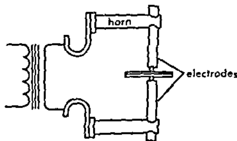


Fig. 1 Spot-welding circuit. When electrodes are closed on the workpiece, the circuit is completed for welding.

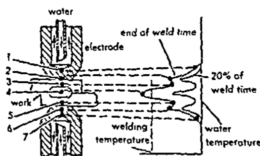


Fig. 2. Distribution of temperature in local (numbered) elements of a spot-welding operation

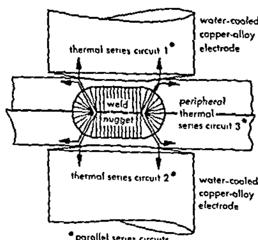


Fig 3. Flow of heat through both electrodes and the workpiece in spot weld is indicated by arrows.

metal that is at fusion temperature at the interfaces

After the metals have been fused together, the electrodes usually remain in place sufficiently long to cool the weld (see Fig 3) An exception is in welding quench-sensitive metals, where it is desirable to remove the electrodes as soon as possible to allow the heat to be conducted in the surrounding metal, preventing steep quench gradients *See* WELDING AND CUTTING OF METALS [E.J L.]

Spot-test analysis

A technique used in qualitative chemical analysis to test for the presence or absence of certain substances Tests are performed by adding a few drops of test reagent to a drop or particle of the unknown substance on a support The unknown is identified by the appearance of a characteristic color reaction The technique is readily adapted to field testing applications It may be used to detect both ionic and molecular substances The advantages of the technique are simplicity speed, and sensitivity. *See* QUALITATIVE CHEMICAL ANALYSIS.

The support is chosen to permit easy detection of color changes during the test. Spot plate, watch glass, ash free filter paper, or spot test paper (a thick and strongly absorbent paper) may be used. The spot plate is best made of colorless transparent glass so that any color change may be observed by placing it on a background of paper of suitably contrasting color.

Substances that may interfere with the test are sometimes separated by a method based upon the different rates of migration of the unknown and the interfering substances through paper. Usually, however, separations are avoided by using sequestering, or masking, agents These convert the interfering substances into precipitates or complexes that do not affect the tests and consequently need not be removed.

The specificity of spot tests is improved by using sequestering agents (which lower the concentrations of the interfering substances) and by controlling the pH of the solutions. [A.A B P.]

Bibliography: F. Feigl, *Spot Tests*, 2 vols., 1954

Spotted fever, Rocky Mountain

An acute, infectious, typhuslike disease of man caused by rickettsialike microorganisms, *Rickettsia rickettsii*, and transmitted by species of ixodid, or hard-shelled, ticks (*see* ACARINA). The disease is now often called American spotted fever because it is found in both North and South America The primary cycle in nature involves chiefly rodents, hares and rabbits, and dogs in some areas plus the opossum and cavy in Brazil, together with the appropriate ticks which infest them The infection in these animals is mostly inapparent. In the labora-

foot pads, and ears. Ticks seem to be the reservoir for the infectious agent. *See* RICKETTSIALES.

In South America, before the identity of the disease had been proven, it was known as São Paulo exanthematic typhus in Brazil and Tobia fever in Colombia Virulence varies in different areas and affects the clinical picture and mortality in man. Incubation in areas with highly virulent strains may be as short as 2–3 days after tick attachment. Onset is sudden with headache, chills, high fever to 105°F, prostration, and appearance in 3–4 days of a measleslike rash The rash, appearing initially on the forearms and ankles, becomes maculopapular over the entire body including the palms and soles; in this respect, the development of the rash differs from epidemic typhus. The temperature may remain high for 2 weeks, falling gradually during the third week if death has not supervened.

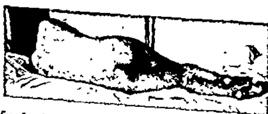


Fig 1. A seve...
(Bitterroot, M...
some necros...
Laboratory) ... Mountain

In milder disease, incubation may be 1–2 weeks and all symptoms are reduced in severity; ambulatory cases have been recognized. According to virulence, mortality has varied in different areas from about 5–80%. Lasting immunity follows recovery, and recrudescences, such as seen in typhus, are not known, though an instance is reported of the isolation of a strain from lymph nodes of a man one year after his recovery. On the contrary, organisms disappear from the tissues of infected guinea pigs, white rats, and ground squirrels in about a month

Rickettsia rickettsii exhibits morphology and staining characteristics similar to *R. prowazekii* They grow in the tick in the cytoplasm and often in the nuclei of all types of cells, including sperm. *In mammals, they occur particularly in cells lining* the walls of blood vessels. Growth in yolk sacs of embryonated chicken eggs is not as rich as with *R. prowazekii*, but is sufficient to provide antigen for vaccines and serologic tests. For over two decades, vaccines were prepared from tissues of laboratory-infected ticks, but at present commercial vaccines are processed from infected yolk sacs Susceptibility to physical and chemical agents, including the broad-spectrum antibiotics, is similar to the typhus organism; hence treatment is also similar. Weil-Felix OX_{19} agglutinins develop during convalescence but cannot be used to differentiate between infections of the typhus and spotted fever groups. *See* PROTEUS; RICKETTSIOSES.

On the other hand, typhus and spotted fever agents are separated by the complement-fixation and agglutination tests. There is confusing cross-fixation, however, with other agents of the spotted fever group, as well as partial to complete cross-immunity in recovered guinea pigs.

In the United States, the tick vectors to man are *Dermacentor andersoni* in the West and southwestern Canada, and *D. variabilis* and *Amblyomma americanum* in the East and South. The rabbit ticks,



Fig. 2. *Rickettsia rickettsii*, causative agent of Rocky Mountain spotted fever in stained smear of infected yolk sac of chicken embryo Extracellular, intracellular, and possibly intranuclear (because of halos around some organisms) forms are depicted. (*Photomicrograph by N. J. Kramis*)

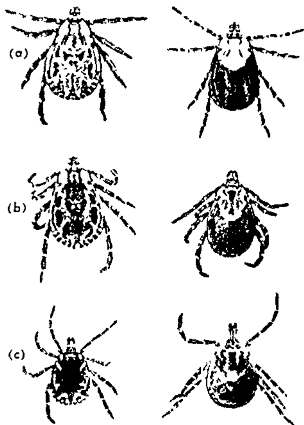


Fig. 3. Vectors of Rocky Mountain spotted fever to man in the United States. (*a*) *Dermacentor andersoni* (*b*) *D. variabilis.* (*c*) *Amblyomma americanum.* Males, left; females, right. (*Photograph by R. A. Cooley from R. L Pullen, Communicable Diseases, Lea and Febiger, 1950*)

Haemaphysalis leporis-palustris over the whole country, and *D parumapertus* in the West, are probably of importance in natural maintenance. In Mexico, *Rhipicephalus sanguineus* and probably *Amblyomma cajennense* carry the infection, and the latter becomes the principal vector in central and northern South America *See* TICK.

All of the above ticks have been found infected in nature in their appropriate areas, but many other species of ixodid ticks can transmit the agent experimentally Natural maintenance of infection occurs in four ways: (1) female ticks pass the organisms to a proportion of their progeny through the eggs; (2) infected, immature ticks feeding on susceptible rodents start new lines of infection in other, simultaneously-feeding ticks; (3) nonhuman tick parasites, such as rabbit ticks, infect rodent hosts, and (4) infected males pass the organism in their sperm to ova during fertilization, and often mate with more than one female [C.B.P.]

Spotted fever group

A term that includes several disease entities in man in various parts of the world, caused by bacterialike organisms, the Rickettsiae The diseases have in common transmission to man by various

species of ticks or certain parasitic mites (*see* ACARINA). The Weil-Felix OX_{19} serology is usually positive, and the pathogens are capable of invading nuclei of host cells. Included in the group are Rocky Mountain spotted fever, fièvre boutonneuse, several tick typhuses, and rickettsialpox. The group of agents is often placed in a distinct subgenus or even separate genus (*Dermacentroxenus*). *See* FIÈVRE BOUTONNEUSE; RICKETTSIALES; RICKETTSIALPOX; SPOTTED FEVER, ROCKY MOUNTAIN; TICK BITE FEVER, SOUTH AFRICAN; TICK TYPHUS, NORTH QUEENSLAND; TICK TYPHUS, SIBERIAN. [C.B.P.]

Spring (mechanical)

A machine element for storing energy as a function of displacement. The flywheel, in contrast, is a means for storing energy as a function of angular velocity. Force applied to a spring member causes it to deflect through a certain displacement thus absorbing energy. A spring may have any shape and may be made from any elastic material. Even fluids can behave as compression springs and actually do in fluid pressure systems. Most mechanical springs take on specific and familiar shapes such as helix, flat, or leaf springs. All mechanical elements behave to some extent as springs because of the elastic properties of engineering materials.

Uses of springs. Energy may be stored in a spring for many uses: to be released later, to be absorbed at the instant the energy first appears, and so on.

Motive power. One of the early and still the most frequent uses of springs is to supply motive power in a mechanism. Common examples are clock and watch springs, toy motors, and valve springs in auto engines. In these, energy is supplied to and stored in the spring by applying a force through a suitable mechanism to deflect or deform the spring. The energy is released from the spring by allowing it to push (as in the valve) or twist (as in a clock) a mechanism through a required displacement.

Return motion. A special case of the spring as a source of motive power is its use for returning displaced mechanisms to their original positions, as in the door closing device, the spring on the cam follower for an open cam, and the spring as a counterbalance. To a certain extent the springs in our vehicles of transportation are in this category. They are designed to keep the car at a certain level with respect to the road or rails, returning the vehicle to this position if displaced by applied forces.

Shock absorbers. Frequently a spring in the form of a block of very elastic material such as rubber absorbs shock in a mechanism. For example, the four legs of a punch press rest on four blocks of rubber. The rubber pads prevent the die-closing inertia forces of the press from transferring down through the legs to the floor with impact or hammer blow proportions. With the rubber pads under the press legs, the force on the floor builds up relatively slowly and no shock is evident. As the acceleration of the die block goes to zero, the inertia force goes to zero and the rubber pad springs, which were deflected by the press blow, are relaxed and ready for the next stroke of the press die. The whole press moves up and down relative to the floor, but by proper selection of the rubber [illegible]

function in vibration control by supplying the necessary flexibility in the support of the vibrating mechanism and the required opposing forces as a result of their deflection. In controlling vibration, the body or mass of the mechanism must be freely supported so as to generate forces opposing the vibrating forces. These opposing forces tend to bring the sum of the forces on the vibrating body to zero. Vibration absorbing mounts are available in a wide range of sizes and spring constants to meet most requirements. These do not prevent vibration, which is a function of speed and the balance of the mechanism, but rather they minimize its effect on the machine frame or the mounting. *See* SHOCK ISOLATION.

Force measurement. Springs have long been used in simple weighing devices. Accurate weighing is usually associated with dead weight devices or balances, but modern spring scales have received wide acceptance and certification for commercial use. Extremely accurate springs for heavy loads are used to calibrate testing machines and in scales over crane hooks for weighing material or devices as they are hoisted. Carefully calibrated

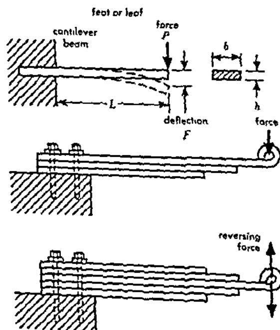


Fig. 1 Flat or leaf spring responds to perpendicular forces in either direction.

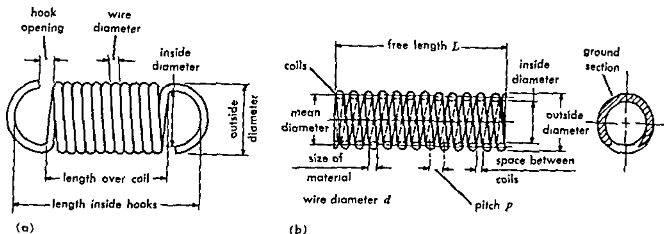


Fig. 2. Helical spring. (a) Wound tight to extend under axial tension. (b) Wound loose to contract under axial compression

springs are used in instruments such as electric meters and pressure gages.

Retaining rings. A relatively modern machine part in which the spring function is used as a holding means is the retaining or snap ring. This device is a split ring of square, rectangular, or special cross section. It fits in a groove on a cylindrical surface or in a bore and stays in place by spring force. A retaining ring prevents or restrains relative axial motion between a shaft or bore and the components on the shaft or in the bore.

Types of springs. Springs may be classified into six major types according to their shape. These are flat or leaf, helical, spiral, torsion bar, disk, and constant force springs.

Flat or leaf spring. A leaf spring is a beam of cantilever design with a deliberately large deflection under a load (Fig. 1). One end of a leaf spring is usually firmly anchored to the frame of the machine and the other end is linked to the moving machine elements by a two-force (pin-ended) link. Force may be tension or compression with no modification of the design. This push-pull feature is the great advantage of a leaf spring, plus the fact that a relatively large amount of energy can be stored in a small space.

Helical spring The helical spring consists essentially of a bar or wire of uniform cross section wound into a helix. The last turn or two at each end of the spring is modified to a plane surface perpendicular to the helix axis, and force can then be applied to put the helix in compression. The ends of the spring helix may be modified into hooks or eyes so that a force may be applied in tension In general it is necessary to design helical springs so that force may be transmitted either in tension or in compression but not both ways for the same spring (Fig. 2). Where reversing forces occur in a spring mechanism, it is better to use a bar or leaf spring and *in some cases a disk spring*

The spiral spring. In a spiral spring, the spring bar or wire is wound in an archimedes spiral in a plane. Each end of the spiral is fastened to the force applying link in the mechanism.

A spiral spring is unique in that it may be deflected in one of two ways or a combination of both of them (Fig 3). If the ends of the spiral are deflected by forces perpendicular to the spiral plane, the spiral is distorted into a conical helix. For better stability and ease of applying the forces, the spring is often made as a conical helix to start with and is deflected into a plane spiral.

A spiral spring may also have the forces act tangent to the spiral as in a clock spring. The

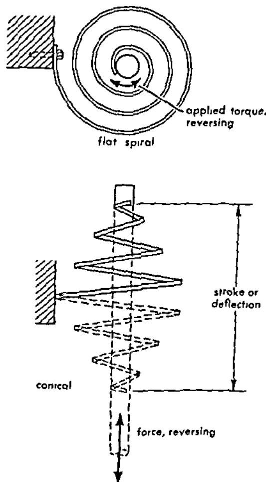


Fig. 3 Spiral spring responds to torsional or translational forces.

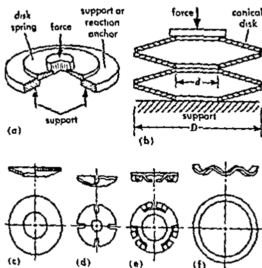


Fig 4 Disk spring occupies small space (a) single tapered disk; (b) multiple disk; (c) Belleville spring; (d) slotted curved washer, (e) finger washer, and (f) wave spring washer

spiral is wound quite open and the tangential force in tension on the spiral tends to close the gaps between successive turns The spring behaves like a beam, bending to a shorter radius of curvature and thus storing energy

Torsion bar A torsion bar spring consists essentially of a shaft or bar of uniform section It stores energy when one end is rotated relative to the other It is used in the spring system of th[illegible] ern motor [illegible]

Disk spr[illegible] space is at a premium the disk spring may be used although it is usually expensive to design and build It consists essentially of a disk or washer supported at the outer periphery by one force (distributed by a suitable chuck or holder) and an opposing force on the center or hub of the disk (Fig 4). For greater deflections several disks may be stacked with the forces transmitted from one to the next by inner and outer [illegible]

Con[illegible] that a [illegible] placem[illegible] counterbalancing of vertically moving masses against the force of gravity is a typical example The Neg'ator spring of Hunter Spring Co provides such a constant force, it uses a tight coil of flat steel spring stock When the outer free end is extended and the coil allowed to rotate on its shaft or pintle, the spring presents a constant restoring force

Spring design. Each type spring has its special features and design refinements Common to all forms of springs are the basic properties of elastic materials Within the elastic range of a material, the ratio of applied force to resulting deflection is constant (see HOOKE'S LAW) Spring systems can be designed to have a variable ratio The ratio is the spring rate or scale and has the dimension of force per unit length.

In a helical spring the elastic action stresses the wire in torsion. The following variables are subject to the designer's action and decision:

free length	L		modules of elasticity in torsion	G
mean diameter	D			
wire diameter	d		number of active turns	N
allowable stress	S		helix angle or coil pitch	P
applied load	P		deflection at load P	F
spring rate	k			

These variables are related by

$$S = 8PD/\pi d^3 \quad (1)$$

which neglects the effect of coil curvature on the stress, and by

$$F = 8PD^3N/Gd^4 \quad (2)$$

For springs subject to frequent cycling as a valve spring, the correction factor for wire curvature must be included The generally used correction is the factor proposed by A. M. Wahl. Introducing this factor changes Eq (1) to

$$S_{max} = (8PD/\pi d^3)\left[\frac{4C-1}{4C-4} + \frac{0.615}{C}\right] \quad (3)$$

in which $C = D/d$ The portion in brackets is [illegible]

[illegible] allowable stress of materials and space

Where compression and tension springs are cycled at very high frequencies, surging may cause high local stress and result in early fatigue failure. Surging is the inability of all parts of the spring to deflect at the same rate due to the inherent inertia in the coils This phenomenon is closely associated with the natural frequency of the spring, and springs should be designed such that their natural frequency and cyclic rate are as far apart as practical [L S L]

Bibliography· American Steel and Wire Division, U S. Steel Company, *Manual of Spring En-*

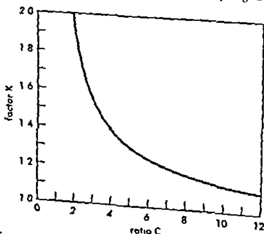


Fig. 5 Correction factor for curvature and shear. (A M. Wahl, Mechanical Springs, Penton Publishing Co, 1944)

gineering; Associated Spring Corporation, *Hand book of Mechanical Spring Design*; K. W. Maier, Your guide to springs that store energy best, *Prod. Eng.*, 71–75, Nov. 10, 1958; A. M. Wahl, *Mechanical Springs*, 1944.

Spring (water)

A place where a concentrated flow of ground water reaches the land surface or discharges into a body of surface water. Springs issue where the water table intersects the land surface. Likewise, where the piezometric surface of an artesian aquifer stands above the land surface, water may issue as a spring if a suitable conduit such as a fault or a solution channel is available through which the ground water can reach the land surface (*see* GROUND WATER).

Springs, including some less-concentrated flows also called seeps, are the chief source of the dry-weather flow of most of our streams. They range between extremely wide limits from mere zones of seepage along river and pond banks to single or multiple orifices that discharge hundreds of millions of gallons of water per day Springs may be classified also by such characteristics as mineral content of the water, temperature, geologic structure, and periodicity of flow.

The largest single spring in the United States is probably Silver Springs in Florida, which has an average flow of more than 700,000,000 gal/day The water issues from limestone into a pool and then flows to the sea in a sizable river. Giant Springs in Montana issues from sandstone and has a discharge of nearly 400,000,000 gal/day Big Spring in Missouri discharges nearly 300,000,000 gal/day from a limestone aquifer. Comal Springs in Texas discharges a similar amount from a limestone aquifer, the water being brought to the surface by a large fault which provides a conduit. Groups of springs such as Malad Springs in Idaho discharge more than 700,000,000 gal/day, and the Thousand Springs in Idaho discharge about 550,000,000 gal/day from lavas of the Snake River Plain [A.N.S.]

Springtail

Any member of the insect order Collembola. The springtails are primitive, wingless insects with no metamorphosis They are mostly less than ⅕ in. long and have chewing mouthparts. They are characterized by a peculiar spring tail, an extension of the ventral surface of the fourth abdominal segment, which is held in place on the underside of the abdomen by a catch. When the spring is released, the little insect is catapulted for some distance through the air. A tube, or collophore, on the ventral surface secretes a sticky substance which enables the springtail to cling to smooth surfaces.

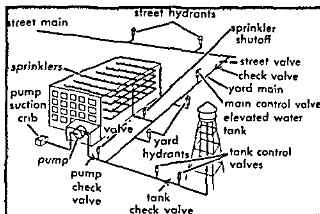


The springtail, *Sminthurus minutus*. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Springtails sometimes occur in great numbers on the surface of stagnant water, in decaying vegetation, and in damp, dark places, such as mushroom houses. Generally of no importance, they may damage mushrooms and, rarely, garden crops. They are easily controlled with rotenone or nicotine sulfate. *See* COLLEMBOLA. [J.D.B.]

Sprinkler, automatic

A safety system used to prevent the spread of fires An automatic sprinkler system is usually installed inside a building as illustrated although sprinklers may also be installed outside for protection against exposure fires.



Typical automatic sprinkler installation. (From F S. Merritt, ed., *Building Construction Handbook*, McGraw-Hill, 1958)

Common automatic sprinkler systems are wet pipe, dry pipe, preaction, and deluge. In a wet pipe system, the pipes contain water and are connected to an ample water supply Individual sprinkler heads are normally closed and are preset to open, usually by the melting of an alloy insert, when the local temperature reaches between 135–165°F for areas having normal atmospheric temperatures Thus, the sprinklers discharge only in the area of a fire and in adjacent areas where hot fumes collect. An alarm may be included that is adjusted to sound when the water flow to the system equals or exceeds the discharge from one sprinkler

In a dry pipe system, the sprinkler pipes contain air under pressure. The opening of a sprinkler allows the air to escape and water to enter the system and to discharge from the sprinkler. Dry pipe systems are used in unheated buildings such as warehouses. In other respects they are similar to wet pipe systems.

In preaction and in deluge systems, separate heat responsive devices in the same area as the

sprinklers admit water to the normally dry sprinkler piping when a fire occurs. The preaction system provides warning before the sprinkler heads operate. In the deluge system, all sprinklers are always open so that water discharges promptly to prevent spread of fire in a hazardous area. The heat responsive device may also sound an alarm.

Local fire codes vary considerably so that a plan for an automatic sprinkler system should be approved by local authorities before construction starts. Insurance rates are usually appreciably lower on buildings having sprinkler systems approved by the underwriter. [F.H.B.]

Spruce

Evergreen trees belonging to the genus *Picea* of the pine family. The needles are single, usually four-sided, and borne on little peglike projections; the cones are *pendulous*. Resin ducts in the wood may be seen with a magnifying lens, but they are fewer than in *Pinus*.

Eastern species. The white spruce, *P. glauca*, ranging from northern New England to the Lake States and Montana and northward into Alaska, is distinguished by the somewhat bluish cast of its needles, small cylindrical cones, and gray or pale brown twigs without pubescence (hairs). Of erect pyramidal habit, it usually attains a height of 60–70 ft with trunk diameters of 1–2 ft, but occasionally in British Columbia and Alberta the tree reaches heights of 80–140 ft with diameters up to 4 ft. When bruised, the leaves of some individuals emit a disagreeable odor, hence the names cat or skunk spruce. The pale, straw-colored wood is soft and straight-grained without different coloration of the heartwood. It is an important source of wood for paper pulp, general construction, and interior finish. The evenly grained wood is much used in the manufacture of musical instruments. The stand in the United States has been estimated at 150,000,000,000 board feet (bd ft), with a much larger volume in Canada. No attempt is made to distinguish the wood of the several species in the trade.

Red spruce, *P. rubens*, is a similar tree but with greener foliage, smaller, more oval cones, and more or less pubescent twigs. Occurring naturally with white spruce in the northeastern United States and adjacent Canada, its range extends southward along the Appalachians into North Carolina. As in white spruce, the greatest use of the wood is for paper pulp. The quantity of eastern spruce pulpwood cut amounts to several times the 100,000,000 bd ft that is annually sawed into lumber. Red spruce makes up about 80% of the volume of approximately 20,000,000,000 bd ft estimated for eastern spruce.

Black spruce, *P. mariana*, ranges from northern New England and Newfoundland to Alaska. However, it occurs sparingly in the Appalachians to West Virginia. The cones are smaller than in the white and red species, are egg-shaped or nearly spherical (1–1½ in. long) and persistent. The twigs are pubescent. Known also as bog or swamp spruce in the southern part of its range, black spruce appears at high elevations in the north. The long-fibered wood is ideal for paper pulp and is also used for paddles, oars, construction, and shipbuilding. It is short-lived as an ornamental tree.

Western species. Blue spruce, *P. pungens*, also known as Colorado blue spruce, is probably the best known of the western species. The twigs are glabrous (without pubescence). The wood is little used, but the tree is popular as an ornamental in northern Europe as well as throughout the United States. Individuals vary in the depth of their blue coloration, the most pronounced being that of the Koster variety. Most of the ornamental specimens come from grafts made on Norway spruce.

Engelmann spruce, *P. engelmanni*, has needles usually of a deep blue-green color, sometimes much like those of the blue spruce, but the young twigs are slightly hairy. The cones, although cylindrical, are smaller than in blue spruce, being about 1½–3 in. long. This species is also a Rocky Mountain tree like the blue spruce, but it is more widely distributed from British Columbia to Arizona and also in the mountains of Oregon and Washington. It is the most important timber spruce of the Rocky Mountain region, the stand having been estimated at about 30,000,000,000 bd ft. In recent years the annual cut has been about 150,000,000 bd ft.

Sitka spruce, *P. sitchensis*, is the largest spruce in the northern hemisphere. The leaves have a pungent odor, are considerably flattened, and stand out from the twig in all directions. Ranging from Alaska to northern California, it occupies a coastal strip about 40–50 miles wide. Mature trees vary from 180–200 ft tall and 3½–6 ft in diameter, but in some instances larger dimensions have been recorded. The stand in the United States is estimated at 5,000,000,000–10,000,000,000 bd ft, nearly all in the states of Washington and Oregon. The annual cut in the United States is about 100,000,000 bd ft. It is used for furniture, doors, blinds, paper pulp, and in piano manufacture for sounding boards.

The Norway spruce, *P. abies*, the common spruce of Europe, is much planted in the United States for



(a,b) Leaves and cone of black spruce, *Picea mariana*. (c) Cone of red spruce, *P. rubens*. (d) Cone of white spruce, *P. glauca*. (e) Cone of Colorado spruce, *P. pungens*. (f) Cone of Norway spruce, *P. abies*. All about ½ natural size. (Brooklyn Botanic Garden)

gineering; Associated Spring Corporation, *Hand book of Mechanical Spring Design;* K. W. Maier, Your guide to springs that store energy best, *Prod. Eng.*, 71–75, Nov. 10, 1958; A. M. Wahl, *Mechanical Springs*, 1944.

Spring (water)

A place where a concentrated flow of ground water reaches the land surface or discharges into a body of surface water. Springs issue where the water table intersects the land surface. Likewise, where the piezometric surface of an artesian aquifer stands above the land surface, water may issue as a spring if a suitable conduit such as a fault or a solution channel is available through which the ground water can reach the land surface (*see* GROUND WATER).

Springs, including some less-concentrated flows also called seeps, are the chief source of the dry-weather flow of most of our streams. They range between extremely wide limits from mere zones of seepage along river and pond banks to single or multiple orifices that discharge hundreds of millions of gallons of water per day. Springs may be classified also by such characteristics as mineral content of the water, temperature, geologic structure, and periodicity of flow.

The largest single spring in the United States is probably Silver Springs in Florida, which has an average flow of more than 700,000,000 gal/day. The water issues from limestone into a pool and then flows to the sea in a sizable river Giant Springs in Montana issues from sandstone and has a discharge of nearly 400,000,000 gal/day. Big Spring in Missouri discharges nearly 300,000,000 gal/day from a limestone aquifer Comal Springs in Texas discharges a similar amount from a limestone aquifer, the water being brought to the surface by a large fault which provides a conduit Groups of springs

lavas of the Snake River Plain [A N.S.]

Springtail

Any member of the insect order Collembola The springtails are primitive, wingless insects with no metamorphosis They are mostly less than ⅕ in. long and have chewing mouthparts They are characterized by a peculiar spring tail, an extension of the ventral surface of the fourth abdominal segment, which is held in place on the underside of the abdomen by a catch. When the spring is released, the little insect is catapulted for some distance through the air. A tube, or collophore, on the ventral surface secretes a sticky substance which enables the springtail to cling to smooth surfaces

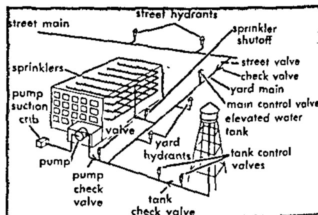


The springtail, *Sminthurus minutus* (*From E. L Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

Springtails sometimes occur in great numbers on the surface of stagnant water, in decaying vegetation, and in damp, dark places, such as mushroom houses. Generally of no importance, they may damage mushrooms and, rarely, garden crops. They are easily controlled with rotenone or nicotine sulfate. *See* COLLEMBOLA. [J D B]

Sprinkler, automatic

A safety system used to prevent the spread of fires. An automatic sprinkler system is usually installed inside a building as illustrated although sprinklers may also be installed outside for protection against exposure fires.



Typical automatic sprinkler installation (*From F. S Merritt, ed, Building Construction Handbook, McGraw-Hill, 1958*)

Common automatic sprinkler systems are wet pipe, dry pipe, preaction, and deluge. In a wet pipe system, the pipes contain water and are connected to an ample water supply. Individual sprinkler heads are normally closed and are preset to open, usually by the melting of an alloy insert, when the local temperature reaches between 135–165°F for areas having normal atmospheric temperatures. Thus, the sprinklers discharge only in the area of a fire and in adjacent areas where hot fumes collect. An alarm may be included that is adjusted to sound when the water flow to the system equals or exceeds the discharge from one sprinkler.

In a dry pipe system, the sprinkler pipes contain air under pressure. The opening of a sprinkler allows the air to escape and water to enter the system and to discharge from the sprinkler. Dry pipe systems are used in unheated buildings such as warehouses. In other respects they are similar to wet pipe systems.

In preaction and in deluge systems, separate heat responsive devices in the same area as the

sprinklers admit water to the normally dry sprinkler piping when a fire occurs. The preaction system provides warning before the sprinkler heads operate. In the deluge system, all sprinklers are always open so that water discharges promptly to prevent spread of fire in a hazardous area. The heat responsive device may also sound an alarm.

Local fire codes vary considerably so that a plan for an automatic sprinkler system should be approved by local authorities before construction starts. Insurance rates are usually appreciably lower on buildings having sprinkler systems approved by the underwriter. [T.M.R.]

Spruce

Evergreen trees belonging to the genus *Picea* of the pine family. The needles are single, usually four-sided, and borne on little peglike projections, the cones are pendulous. Resin ducts in the wood may be seen with a magnifying lens, but they are fewer than in *Pinus*.

Eastern species. The white spruce, *P. glauca*, ranging from northern New England to the Lake States and Montana and northward into Alaska, is distinguished by the somewhat bluish cast of its needles, small cylindrical cones, and gray or pale brown twigs without pubescence (hairs). Of erect pyramidal habit, it usually attains a height of 60–70 ft with trunk diameters of 1–2 ft, but occasionally in British Columbia and Alberta the tree reaches heights of 80–140 ft with diameters up to 4 ft. When bruised, the leaves of some individuals

cat or

is soft a

tion of

wood for paper pulp, general construction, and interior finish. The evenly grained wood is much used in the manufacture of musical instruments. The

wood of the several species in the trade

Red spruce, *P. rubens*, is a similar tree but with greener foliage, smaller, more oval cones, and more or less pubescent twigs. Occurring naturally with white spruce in the northeastern United States and adjacent Canada, its range extends southward along the Appalachians into North Carolina. As in white spruce, the greatest use of the wood is for paper pulp

amour

is ann

up ab

000 000

Alaska. However, it occurs sparingly in the Appalachians to West Virginia. The cones are smaller than in the white and red species, are egg shaped or nearly spherical (1–1¼ in. long) and persistent. The twigs are pubescent. Known also as bog or swamp spruce in the southern part of its range, black spruce appears at high elevations in the north. The long-fibered wood is ideal for paper pulp and is also used for paddles, oars, construction, and shipbuilding. It is short lived as an ornamental tree.

Western species. Blue spruce, *P. pungens*, also known as Colorado blue spruce, is probably the best known of the western species. The twigs are glabrous (without pubescence). The wood is little used, but the tree is popular as an ornamental in northern Europe as well as throughout the United States. Individuals vary in the depth of their blue coloration, the most pronounced being that of the Koster variety. Most of the ornamental specimens come from grafts made on Norway spruce.

Engelmann spruce, *P. engelmanni*, has needles usually of a deep blue-green color, sometimes much like those of the blue spruce, but the young twigs are slightly hairy. The cones, although cylindrical, are smaller than in blue spruce, being about 1½–3 in. long. This species is also a Rocky Mountain tree like the blue spruce, but it is more widely distributed from British Columbia to Arizona and also in the mountains of Oregon and Washington. It is the most important timber spruce of the Rocky Mountain region, the stand having been estimated at about 30,000,000,000 bd ft. In recent years the annual cut has been about 150,000,000 bd ft.

Sitka spruce, *P. sitchensis*, is the largest spruce in the northern hemisphere. The leaves have a pungent odor, are considerably flattened, and stand out from the twig in all directions. Ranging from Alaska to northern California, it occupies a coastal strip about 40–50 miles wide. Mature trees vary from 180–200 ft tall and 3½–6 ft in diameter, but in some instances larger dimensions have been recorded. The stand in the United States is estimated at 5,000,000,000–10,000,000,000 bd ft, nearly all in the states of Washington and Oregon. The annual cut in the United States is about 100,000,000 bd ft. It is used for furniture, doors, blinds, paper pulp, and in piano manufacture for sounding boards.

The Norway spruce, *P. abies*, the common spruce of Europe, is much planted in the United States for



(a,b) Leaves and cone of black spruce, *Picea mariana*. (c) Cone of red spruce, *P. rubens*. (d) Cone of white spruce, *P. glauca*. (e) Cone of Colorado spruce, *P. pungens*. (f) Cone of Norway spruce, *P. abies*. All about ⅓ natural size. (Brooklyn Botanic Garden)

timber as well as for ornamental purposes. It can be recognized by the dark green color of the leaves, by glabrous, pendent, short branchlets, and by cones 4 6 in. in length, usually near the top of the tree. An important source of paper pulp, it has often been planted for that purpose in the northern and eastern United States. It is also used for shelter belts and for Christmas trees, although not as desirable for the latter purpose as the firs and Douglas-fir, since it sheds its needles quickly in a warm room. There are many cultivated varieties *See* FOREST AND FORESTRY; TREE [A H.G.]

Sputtering

The process by which atoms or groups of atoms are *ejected from a metal surface as the result of heavy-ion impact It generally takes place at the cathode* of a self-maintained gaseous discharge, indicating that the important agent is the positive ion. Although sputtering is useful for certain processes, such as the generation of a clean surface, it is usually harmful. In the case of an oxide-coated thermionic cathode, sputtering by positive-ion bombardment may destroy the surface completely.

In general, there will be a threshold energy for sputtering. This depends on both the surface material and the bombarding ion It has been found empirically that the mass m sputtered per unit time is given by

$$m = k\,(V_c - V_0)$$

where k and V_0 are constants, while V_c is the potential through which the ion has fallen, and is *therefore a measure of the energy of the ion* Further, it has been shown by A. Guntherschulze that the sputtered mass decreases as the pressure increases. The sputtering process becomes more efficient as the mass of the ion is increased The ejected atoms may come off either as neutral particles or as ions.

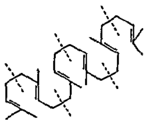
The explanation of sputtering is not clear It is thought that local heating may play a prominent part in the process *See* CANAL RAYS; METAL COATINGS. [C H MI]

Bibliography: L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases*, 1939.

Squalene

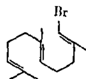
A triterpene which appears to play a role in the biosynthesis of sterols and polycyclic terpenes.

The dotted lines indicate six isoprene units from which the $C_{30}H_{50}$ hydrocarbon is theoretically



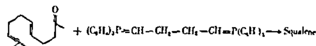
Squalene

condensed. Actually it has been synthesized by the condensation of two molecules of farnesyl bromide



Farnesyl bromide

and more recently by applying the Wittig reaction to a pure *trans*-geranylacetone



The squalene so formed was probably a mixture of three stereoisomers. Later work showed that purification could be effected by forming a thiourea clathrate. There is an indication that synthetic squalene can be enzymatically converted to lanosterol.

The six olefinic bonds in squalene have been demonstrated to be of the trans configuration. It possesses a faint, pleasant odor and a high boiling point. Decomposition occurs when distilled at ordinary pressures. When exposed to air, squalene absorbs oxygen and resinifies to a viscous mass. It is insoluble in water but soluble in fats and fat solvents.

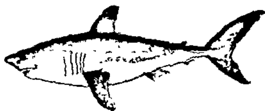
Squalene is found in appreciable quantities in *the liver of sharks Smaller amounts (0.1–0.8%)* occur in olive oil, wheat germ oil, rice bran oil, and yeast, as well as in human sebum and ear wax *See* ISOPRENE; TERPENE, TRITERPENE. [F L.S.]

Squaliformes

The Squaliformes, or sharks, constitute one of the Recent orders of the subclass Elasmobranchii Sharks also known as the Selachii, are distinguished from skates and rays (order Rajiformes) in having lateral gill slits, a pectoral fin which is free anteriorly, and the upper margin of the orbit free from the eyeball Sharks date from the Devonian, but most modern types have evolved since the late Mesozoic.

Modern sharks may be arranged in 7 suborders, 14 families, roughly 80 genera, and about 225 species. Sharks occur in all oceans, mostly in shore and surface waters of tropical to temperate seas Some forms, a few of which have luminescent organs, descend to depths of 1500 fathoms, and others invade the lower channels of major rivers (*see* Bioluminescence). One species, possibly different from its marine relatives, occurs in Lake Nicaragua

All sharks are carnivorous. The sluggish whale shark and the basking shark, largest of all fishes feed on tiny organisms. Many species, including especially the white shark, have justifiably sinister reputations because of attacks on swimmers.



Salmon shark, *Lamna ditropis*. (After G. B. Goode, Great International Fisheries Exhibition, London, 1883, U.S. Natl. Museum Bull. 27, 1884)

Internal fertilization is accomplished with pelvic-fin claspers. Some species lay eggs with horny, rectangular cases, others retain encased eggs in the body and bear fully formed young; and still others are truly viviparous, the young being fed through food transfer between the uterine wall and the yolk sac.

Shark fisheries are generally of limited importance except locally. Some sharks yield a vitamin-rich liver oil, the skin of large individuals offers a valuable source of leather, and the fins are esteemed as an item of diet in the Orient. On the other hand, certain sharks may damage fishing gear or destroy valuable food fish already caught in nets or on lines. See ELASMOBRANCHII. [R.M.B.]

Bibliography. H. B. Bigelow and W. C. Schroeder, Sharks, *Fishes of the Western North Atlantic*, Sears Foundation Marine Research, Mem. 1, pt. 1, 1948.

Squall

A strong wind with sudden onset and more gradual decline, lasting for several minutes. Wind speeds in squalls commonly reach 30–60 mph, with a succession of brief gusts of 80–100 mph in the more violent squalls. Squalls may be local in nature, as with isolated thunderstorms, or may occur over a wide area in the vicinity of a well-developed cyclone, where the squalls locally reinforce already strong winds. Because of their sudden violent onset and the heavy rain, snow, or hail showers which often accompany them, squalls cause heavy damage to structures and crops, and present severe hazards to transportation.

The most common type of squall is the thundersquall or rainsquall associated with heavy convective clouds, frequently of the cumulonimbus type. Such a squall usually sets in shortly before onset of the thunderstorm rain, blowing outward from the storm and generally lasting for only a short time. It is formed when cold air, descending in the core of the thunderstorm rain area, reaches the earth's surface and spreads out (see THUNDERSTORM). Particularly in desert areas, the thunderstorm rain may largely or wholly evaporate before reaching the ground, and the squall may be dry, often associated with dust storms (see DUST STORM).

Squalls of a different type result from cold air drainage down steep slopes. The force of the squall is derived from gravity, and depends on the descending air which is colder and more dense than the air it replaces. So-called fall winds of this kind are common on mountainous coasts of high latitudes, where cold air forms on elevated plateaus and drains down fiords or deep valleys. Channeling of the air through narrow valleys increases their force. Squall winds (bora) of 100 mph or more are observed in winter, when arctic air masses from Russia intermittently spill over the mountains of Yugoslavia into the relatively warm coastal regions of the Adriatic.

Squalls also occur over a wide area on passage of a cold front, where weaker general winds in advance of the front are suddenly replaced by stronger winds in the cold air. In this case, high gusty winds may be sustained for many hours.

A line along which a number of squalls occur in association with thunderstorms is designated a squall line. These lines, which may be several hundred miles long, consist of several large, aligned clusters of thunderstorms, 30–80 miles across, with weaker activity between the active thunderstorms. Passage of the active parts is marked by sudden strong gusty winds, usually with pronounced veering of the wind direction, a rapid drop of temperature (10–15° in a few minutes), onset of heavy rain, thunder, lightning, and often hail. After the first few minutes, a gradual decline in wind speeds and intensity of weather phenomena is observed; the rain is usually over in about 45 min, by which time the wind often returns to the direction and strength present before squall line passage. A sudden jump in atmospheric pressure, amounting to 2–5 millibars, is also observed, followed by a gradual decline. Between the active parts, only weak squalls and weather conditions may be found. The active portions, where turbulent convective clouds rise to 10–13 km, present a severe hazard to aircraft, but may be circumnavigated with the aid of ground and air-borne radar.

Squall lines generally form at or near cold fronts of extratropical cyclones. The thunderstorms of which they consist extend up into regions of strong winds well above the earth's surface, and being generally blown away from the cold front, are most often found 100–400 km in advance of the front. Formation requires unstable moist air, and thus squall lines in the United States are most common in spring and early summer when northward incursions of maritime tropical air masses take place east of the Rockies and interact with polar front cyclones. Most tornadoes occur under squall line conditions. [C.W.N.]

Bibliography. T. F. Malone (ed.), *Compendium of Meteorology*, 1951.

Squamata

The dominant order of living reptiles composed of the lizards and snakes. The group first appeared in Jurassic times and today is found in all but the coldest regions. Various forms are adapted for ar-

boreal, burrowing, or aquatic lives but most squamates are fundamentally terrestrial. There are about 4700 recent species, 2200 lizards and 2500 snakes.

The order is readily distinguished from all known reptiles by its highly modified skull. The immediate ancestors of primitive lizards are members of the diapsid order Eosuchia in which a pair of temporal foramina are present on both sides of the head. A similar pattern is found among living reptiles in the tuatara, order Rhynchocephalia, and the crocodilians, order Crocodilia. In the Squamata, however, in the case of the quadratojugal and jugal bones which border the lower margin of the inferior temporal opening, the former has been lost completely and the latter greatly reduced. In consequence there is only a single temporal opening present and the quadrate has become enlarged and movable. Even this temporal opening is lost or reduced in many forms. No other reptiles show these modifications which allow for great kinesis in the lower jaw since it articulates with the quadrate. In addition the order is distinct from other living reptile groups because its members have no shells, no secondary palates, and the males possess paired penes.

Lizards versus snakes. Traditionally the Squamata have been divided into two major subgroups, the lizards, suborder Sauria, and the snakes, suborder Serpentes. The latter group is basically a series of limbless lizards and it is certain that snakes are derived from some saurian ancestor. It is by no means certain, however, that the snakes are a natural group. There are many different legless lizards and it has been suggested that more than one line has evolved to produce those species currently grouped together as snakes. Besides this possibility there is the additional problem of separating the two suborders. No single attribute will suffice for this purpose but the features in the following table, used in combination, will definitely allocate any questionable form to one or the other group.

Comparison of Sauria and Serpentes

Determining structure	Sauria	Serpentes
Limbs	*Usually 2 or 4, sometimes none*	*None*
Eyelids	Usually movable, sometimes immovable	Immovable
External ear opening	*Usually present, sometimes lacking*	*None*
Pectoral girdle	Usually present, rarely none	None
Braincase	*Usually open anteriorly, bounded by connective tissue sheath instead of bone*	*Completely bony anteriorly*
Mandible	Two halves usually articulating	Two halves connected by elastic ligament, not articulating

SAURIA

The majority of saurians are insectivorous but a few feed on plants while others, most notably the Varanidae and allies, feed on larger prey including birds and mammals. The largest living lizard is the Komodo dragon, *Varanus komodoensis*, which attains a length of 10 ft and a weight of 50 lb. Several small geckos and African chameleons are about 1½ in. long when fully grown.

Physiology. The senses of lizards include the usual ones found in all terrestrial vertebrates. Thermal, tactile, and pain receptors are distributed over the body surface.

Olfaction. The principal chemoreceptors are the nasal organ proper, and a specialized derivative of the olfactory apparatus, the organ of Jacobson or the vomeronasal organ. The nasal organs open into the mouth cavity through the internal nostrils and are connected to the outside by the external nostrils. Inspired air passes through the organ on its way to the throat and lungs just as in all other terrestrial vertebrates. Small particles of substances in the air are trapped by the nasal mucosa and analyzed by sense cells in this lining. The paired organs of Jacobson lie anterior to the olfactory

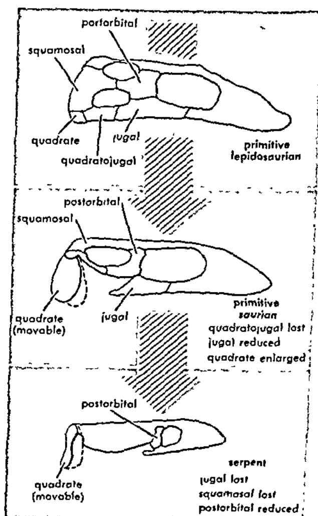


Fig. 1. Diagrams of skulls in various reptile groups, indicating evolution of squamate skull types.

sacs and are separate from them. They open only into the mouth and particles in watery solution from this region are analyzed by them. In many lizards the *tongue* has a divided tip and the two divisions are used to pick up particles outside the mouth and bring them back into it. The tips are then rubbed against the openings of the vomeronasal organs Jacobson's organs are indicated as separate areas of the nasal organ in Amphibia and turtles, are partially separated from the main sac in Rhynchocephalia, and are vestigial in crocodiles, birds, and most mammals. *See* CHEMORECEPTION; SENSE ORGAN.

Hearing. The auditory apparatus of lizards shows remarkable variation correlated with the mode of life In the majority of forms a well-developed *eardrum* is present in a slight recess at the posterior margin of the head near the quadrate bone. A ligament attaches the small middle-ear ossicle, the stapes, to the inner face of the tympanum and the other end fits into an opening into the inner-ear region. Sound is transmitted from the eardrum along the stapes to the fluid of the inner ear In many burrowing lizards the eardrum is partially or completely covered by skin and scales or is lost entirely Most of these forms are insensitive to airborne sounds, and receive their auditory stimuli from the substratum through the lower jaw to the quadrate and then to the stapes.

Vision Lizards have moderately good vision except for burrowing forms in which the eyes are reduced or covered over by skin, scales, or even bone Most lizards with unreduced eyes have color vision, but some nocturnal families such as the Gekkonidae and Xantusiidae are color-blind In certain forms the movable eyelids have a transparent window through which an image may be seen even when the eyes are closed In several groups the two eyelids have fused and are completely transparent. The fused lids form a protective spectacle in geckos, night lizards or Xantusiidae, certain burrowing skinks and other forms.

Locomotion The majority of lizards are quadrupedal in locomotion, and are usually ambulatory scamperers or scansorial. Some forms are bipedal, at least when in haste. Perhaps the most famous of the bipedal forms is the *Jesus Cristo* lizard, *Basiliscus*, of Middle America which runs for some distance across the surface of small streams when frightened In many burrowing lizards the limbs are reduced so that sometimes only the anterior pair is present or all are gone In these species locomotion is of a serpentine curvilinear type One genus of Asian lizards, *Draco*, has large membranes between its limbs that can be used as gliding surfaces to increase the distances of leaps and to reduce the rate of descent

Thermoregulation Lizards, in common with all other reptiles, are dependent upon external sources for maintenance of their body temperatures. They are ectothermic. Thermoregulation is obtained by means of gross behavioral adjustments. A lizard moves into shade if its temperature rises and moves into areas of higher temperatures as its temperature drops. Individuals are most active between temperatures of 60°–108°F with optimum activity between 80–90°F. Temperatures above 110° are lethal and no lizard can survive in direct desert sunlight for more than 15–30 minutes without shelter. *See* THERMOREGULATION.

Respiration and excretion. Respiration is achieved by means of paired lungs. The excretory product of protein metabolism is an insoluble substance, uric acid The feces are dark, oblong, and capped by a white mass of uric acid when deposited.

Coloration. The coloration of each species of lizard is characteristic. Most forms exhibit marked differences in coloration between the sexes, at least during the breeding season, and frequently the young are markedly different from the parents. Color changes occur in rapid fashion among some species and all are capable of metachrosis or changing color to a certain extent. Contrary to popular legend the changes are not made to match the color of the background but seem to be under neurohormonal control and fluctuate with temperature, light, and the activities of the lizard. *See* CHROMATOPHORE; SEXUAL DIMORPHISM.

Sound production Many lizards make sounds ranging from hisses produced by simple expulsion of air through the glottis to squeaks of small forms, and loud cluckings of the geckos. Sounds are used for defensive bluffing and probably for recognition as well

Defense mechanisms. The basic defense against discovery utilized by most lizards is "freezing" motionless in position The majority are protectively colored and when still are extremely difficult to see However, if located, rapid flight becomes the second line of defense. Even if a predator catches a lizard the game is not over, for many forms bite and claw Among the more interesting specialized defensive devices of saurians are the bluffing techniques of certain forms The Australian frilled lizard *Chlamydosaurus*, of the family Agamidae, can expand a huge nuchal collar which when presented with open mouth to a predator is an effective deterrent to aggression. The agamid, *Uromastix*, of Africa and the iguanid, *Ctenosaura*, of Middle America have spiny tails that may lash out to injure a potential enemy The horned lizards, *Phrynosoma*, of the United States and Mexico, are armed with numerous spines, particularly on the head, and would make a rather [illegible] [illegible] repugnant to coyotes and dogs. A great many lizards are able to throw off their tails voluntarily (tail autotomy) and reflexes in the tail make it thrash about to distract a predator from its real objective In many cases the tail is brightly colored to add to the illusion. On the lizard the tail will regenerate, but the new tail will lack a bony support *See* AUTOTOMY

Venomous species. There are but two species of venomous lizards, both members of the genus *Heloderma*, in the family Helodermatidae. *H. suspectum*, the Gila monster, is found in western New Mexico, Arizona, extreme southern Nevada, and northwest Mexico. The beaded lizard, *H. horridum*, is a Mexican species. Venom is produced in glands along the lower jaw and penetrates wounds inflicted by the anterior recurved teeth. No mechanism for injection of the venom into the wound is present, contrary to the situation in most snakes. Although occasionally fatal, the bites of these large lizards, of which a length of 3 ft is not uncommon, are not as dangerous as those of snakes, nor are human beings bitten frequently.

Rhythmic activity. Saurians show rather definite rhythms in their activities. In temperate regions seasonal cycles are correlated with temperature and sexual activity, and hibernation occurs in the winter. General activity continues the year around in tropic climates. Daily activity cycles are typical of most species, and include a definite daily pattern of feeding, drinking, resting, basking, and excreting. Most lizards are diurnal and their diel cycles are usually correlated with temperature. Certain families, notably the Gekkonidae and Xantusiidae, tend strongly toward nocturnal habits.

Territoriality. Most lizards spend their entire life within a rather restricted area, and return at the end of each day to the same home site to sleep. The males of the majority of forms exhibit territorial behavior and will defend a central portion of their home range against trespass by other males of the same species. Defense of the territory usually involves much bluffing and noise, and the "defending" lizard is usually the victor. Various color patches, scale crests, and other distinctive marks play a part in intimidation and display at these times. Fights frequently occur and generally consist of attempts at biting. Male monitor lizards, Varanidae, rear up on their hind legs and bite and slash with the front claws. No truly gregarious lizards are known and none are migratory. See TERRITORIALITY.

Courtship activity. Males usually breed with any female that they meet in moving about their home ranges during the breeding season. A complex courtship pattern involving a chase and considerable prenuptial nuzzling and biting is typical. During copulation the male frequently grasps the head or neck of the female in his mouth and the tails of the breeding pair are entwined so that their cloacas are brought into contact. The male has paired hollow hemipenes which normally lie retracted into the base of the tail. When engorged with blood these organs are everted to be used singly in the mating act.

Reproduction. Most species produce eggs that have leathery or calcareous shells. The eggs are buried in the soil, hidden in decaying logs, or under bark. In many cases the female guards the eggs against predators. Some species are ovoviviparous and retain the eggs within the body until they hatch. Others are truly viviparous with a sero-allantoic placental connection between mother and embryo. Gestation or incubation is generally 5–8 weeks. In oviparous and ovoviviparous forms, at least, the young have a special egg tooth which grows up from the tip of the upper jaw. This tooth is used to cut through the egg membranes and shell at hatching and is lost shortly thereafter.

Classification of lizards. The following list indicates the major evolutionary lines, families, and distribution of lizards. Families indicated by an asterisk (*) contain limbless snakelike species. All members of the families Pygopodidae, Dibamidae, Amphisbaenidae and Anniellidae are snakelike lizards.

Iguania line
- Family Iguanidae: the Americas, Madagascar, Fiji
- Family Agamidae: Africa, southern and central Asia, Australia
- Family Chamaeleontidae: Africa, Madagascar, south India

Gekkota line
- Family Gekkonidae: circumtropical, all continents and most continental and oceanic islands
- Family Pygopodidae: Australia

Scincomorpha line
- Family Xantusiidae: North America, Cuba
- Family Teiidae: the Americas
- Family Lacertidae: Africa, Eurasia
- *Family Gerrhosauridae. Africa, Madagascar
- *Family Scincidae: cosmopolitan, except frigid areas
- *Family Dibamidae: Indo-Malaysia, Philippines, and New Guinea
- *Family Amphisbaenidae: tropics of Africa and America

Anguimorpha line
- *Family Anguidae: North and Middle America, Eurasia
- *Family Anniellidae: the Californias
- Family Xenosauridae: southeast China, Central America
- Family Helodermatidae: North America
- Family Varanidae. Africa, southern Asia, Australia
- Family Lanthonotidae: North Borneo

SERPENTES

Morphology. Snakes are basically specialized limbless lizards which were probably evolved from burrowing forms but have now returned from subterranean habitats to occupy terrestrial, arboreal and aquatic situations. In addition to those features already mentioned as distinguishing the snakes from lizards the following characteristics are typical of all serpents. There is no temporal arch so that the lower jaw and quadrate are very loosely attached to the skull. This gives the jaw even greater motility than is the case in lizards. The

body is elongate with 100–200 or more vertebrae and the internal organs are elongate and reduced A spectacle covers the eye and there is no tail autotomy. In most features of their biology the Serpentes resemble their close relatives, the lizards.

The largest living snake is the Indian python, *Python reticulatus*, which reaches 30 ft in length and a weight of 250 lb. The largest venomous snake is the king cobra, *Ophiophagus hannah*, of southern Asia, which is known to attain a length of 18 ft A number of small tropical burrowing snakes are midgets by comparison and are about 5 in. in length when fully grown. The longevity of snakes is not well known but one captive cobra has lived 28 years and many of the larger snakes probably live at least 25 years

Physiology. The senses of snakes are fundamentally similar to those of the saurians Great dependence is placed upon olfaction and the Jacobson's organs. The tongue of all snakes is elongate and deeply bifurcated When not in use it can be retracted into a sheath located just anterior to the glottis but it is protrusible and is constantly being projected to pick up samples for the Jacobson's organs from the surrounding environment Snakes are deaf to air borne sounds and receive auditory stimuli only through the substratum via the bones of the head The eyes are greatly modified from those in lizards and there is no color vision Some groups are totally blind and have vestigial eyes covered by scales or skin.

Thermoreception Specialized thermoreceptors, not found in any other vertebrates, are present in many boas and pythons, family Boidae, and all pit vipers Crotalidae In the boids these receptors form a series of de...

... to temperature differences of about 2°F and are apparently used to locate prey whose temperature may be higher or lower than that of the surrounding environment

Locomotion Four basic patterns of locomotion are found in snakes and several may be used by a particular individual at different times The most familiar type is serpentine curvilinear In this pattern the snake moves forward by throwing out lateral undulations of the body and pushing them against any irregularity in the surface. Snakes using rectilinear locomotion move forward in a straight line without any lateral undulations, by producing wavelike movements in the belly plates Laterolinear locomotion or sidewinding is used primarily on smooth or yielding surfaces and is

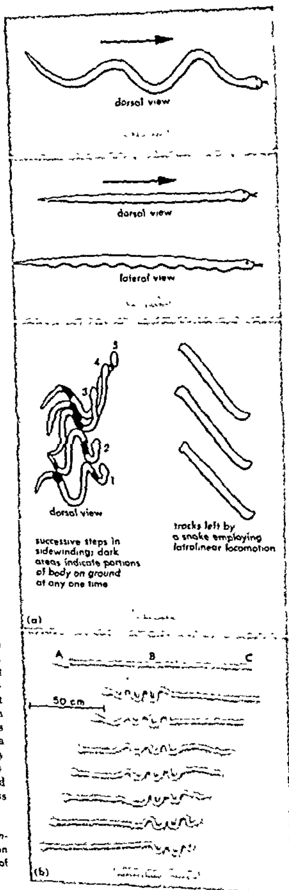


Fig 2 Squamata (a) Locomotion in snakes (b) Concertina locomotion. Successive stages showing extension of head and neck followed by pulling forward of body and tail, coils anchor body

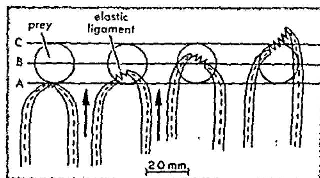


Fig 3. (1–4) Dorsal view of lower jaw, demonstrating how a snake swallows prey.

very complex. In essence, the snake anchors a portion of the body in the substratum and lifts the rest out laterally to a new position When the lifted part is anchored again, the portions of the body left behind are lifted to the new position. By a constant lifting and anchoring of alternate parts the snake moves in a lateral direction. Concertina locomotion movement resembles the expansion and contraction of that musical instrument (see Fig 2).

Other activities. The basic bodily functions of snakes—thermoregulation, respiration, excretion, and elimination—are the same as those in lizards.

Defense mechanisms. Snakes and lizards are alike in utilizing procrypsis and flight as their principal means of defense Protective coloration is of prime importance in the first of these responses Although all snakes are capable of some color change, none show the marked changes characteristic of many lizards. Snakes are excellent bluffers, and hissing and striking, sometimes with mouth closed, may be used to intimidate or discourage possible predators. Among the most effective deterrents to attack are the specialized warning devices of the rattlesnakes from which they receive their names. Rattlesnakes, however, are not bluffing, because they can back up their threats with an injection of venom.

Ecology. The seasonal and daily activity patterns of snakes are similar to those described above for lizards More snakes are nocturnal than diurnal and they generally have a larger home range than do lizards. Home sites are found in most species but no territorial behavior is known. A "combat dance" in which the males rear upright and push and weave around one another occurs frequently The dance has often been described as being a courtship display, but it always involves two males. Its exact function remains a mystery. Snakes are generally nongregarious although the sea snakes, Hydrophiidae, travel in large schools. Migrations occur in temperate regions to and from denning sites and the sea snakes migrate annually to breeding grounds.

Courtship. Courtship is distinctive for each form, but usually involves tracking of the female by the male, considerable nuzzling and rubbing of the female by the male, and a wrapping of the bodies of the breeding pair tightly around one another for mating. Male snakes utilize their hemipenes one at a time Eggs are buried or hidden, and are leathery or calcareous Development is oviparous, ovoviviparous, or viviparous Incubation or gestation takes about 60–100 days

Nutrition. Food is almost always of animal material such as amphibians, reptiles, birds and their eggs, and mammals However, some forms eat lizard and snake eggs to a large extent, insects, mollusks, or fish, and most snakes probably eat carrion The majority of species are rather specific in food preference. Prey may simply be grabbed and swallowed, or specialized techniques may be employed.

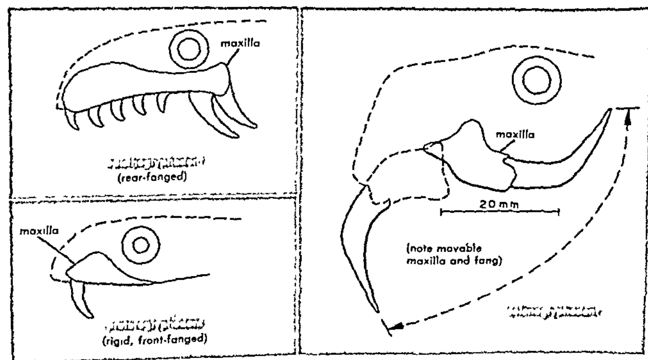


Fig. 4 Dentition of venomous snakes

depending upon the species. Constriction of the victim is practiced by many forms. In this case the food organism is large and is grasped in the snake's mouth while coils of the muscular body are wrapped around the prey The coils are gradually tightened until the victim's heart stops or it suffocates. Venom may be injected into the body of the prey by certain forms. Food articles are always swallowed whole The process is complex but consists primarily of a series of coordinated movements of the head and jaws The lower jaws are connected to one another by an elastic ligament which allows each half to be moved independently of the other The teeth on the jaws are recurved so that by alternately moving the right half of the jaw forward and then the left and hooking the teeth into the prey the snake literally pulls itself forward over the prey's body

Venomous snakes. The vast majority of living snakes are harmless to man, although a number are capable of inflicting serious injury with their venomous bites The venom apparatus has evolved principally as a method of obtaining food, but it is also advantageous as a defense against attackers. Venomous snakes may be placed in three categories on the basis of their venom apparatus These categories are termed opisthoglyphodont in those snakes which have fangs in the rear of the mouth Proglyphodont snakes have fangs in the front of the mouth This group can be subdivided into the proteroglyphodont in which the anterior fangs are fixed and immovable, for example, cobras, mambas, kraits, coral snakes, and sea snakes, and the solenoglyphodont in which the fangs are inserted on the movable maxilla so that they may be rotated forward when the mouth is open and folded back against the roof of the mouth when not in use as in the vipers

Fangs are teeth modified for the injection of venom into the victim, and the venom glands are modified salivary glands connected to the grooved fangs by a duct. Special muscles are present in all proglyphous snakes to force the venom into the wound The venom itself is a complex substance containing a number of enzymes Certain of these enzymes attack the blood, others the nervous system, and some are spreaders The venom of each species is distinct and contains different kinds, combinations, and proportions of the deleterious enzymes

The rear-fanged and proteroglyphodont forms have relatively short fangs and require some chewing of the victim in order to inject the venom Small rear-fanged forms are consequently harmless to human beings because they cannot get their mouths around any part of a person's body to bring the fangs into contact. Vipers, however, have very long fangs and they usually strike or bite with lightning speed Their injection mechanism is efficient and a bite usually means that a considerable amount of venom has been injected into the victim

Classification of snakes. The following list indicates the major groups of living snakes and their distribution. The family, Typhlopidae, is a questionable snake group because its members may be limbless lizards.

... circumtropical
... al
... and Central America

Family Boidae: circumtropical, western United States

Family Colubridae: cosmopolitan, except frigid areas; some species are opisthoglyphodont

Family Elapidae: Africa, Asia, Australia, the Americas; all species are proglyphodont

Family Hydrophiidae: oceans of the Indo-Pacific region, from Africa to off tropical America; all species are proglyphodont

Family Viperidae: Eurasia, Africa; all species are proglyphodont

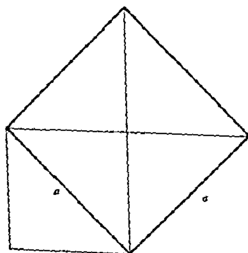
Family Crotalidae: southern and eastern Asia, the Americas; all species are proglyphodont

[J.M.S.]

Bibliography: J. A. Oliver, *The Natural History ... Reptiles*, 1955; ... *Life*,

Square

A portion of a plane bounded by four equal line segments, or sides, each two of which are either parallel or mutually perpendicular The intersecting sides determine four points, vertices of the square. The diagonals join pairs of vertices that are not on a side They are equal and mutually perpendicular. If a is the length of a side of a square, its area is $a \times a$ or a^2, and this is the motivation for calling a^2, "a squared." The square is symmetric about its two diagonals and the two lines that join the midpoints of opposite sides. The square of unit side is the unit of area Perhaps the most famous theorem of Greek geometry concerning the square is the theorem of Pythagoras which asserts that the square erected on the hypote-



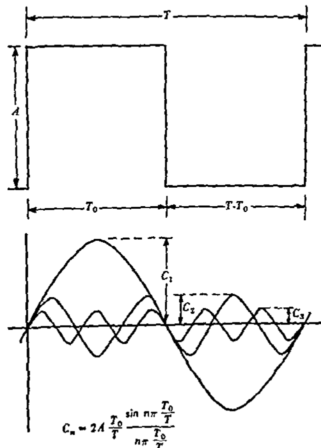
Square with side a

nuse of a right triangle has an area that is the sum of the areas of the squares erected on the two legs of the triangle. Special cases of the theorem were known to the ancient Egyptians, who applied it in surveying. There is reason to doubt that a rigorous proof of the theorem was obtained by Pythagoras or his school. The proof of the theorem given in Euclid's *Elements* (Book I, Proposition XLVII) is generally credited to Euclid himself. *See* PYTHAGOREAN THEOREM; QUADRILATERAL; RECTANGLE.
[L.M.BL.]

Square wave

A recurrent time-varying function that maintains a constant value for a fixed interval, shifts to another constant value for another fixed interval, and repeats these alternations at equally recurring periods. When the two portions of the wave do not occupy equal time intervals, the function is sometimes referred to as a rectangular wave.

Analysis. For purposes of analysis, square waves may be expressed mathematically by a Fourier series consisting of the sum of harmonically related sinusoids, the lowest-frequency component having a period equal to the period of the square wave. For example, the symmetrical waveform in the illustration can be approximated by the sum of the fundamental and odd harmonics having a specific phase relationship. The broad, flat top may be fairly well reproduced by the use of the fundamental and a few of the lower-order harmonics, whereas accurate representation of the sharp transitions requires a larger number of the higher-frequency components. A detailed Fourier analysis would show that as the number of terms in the series approaches infinity, true reproduction of the square wave would be approximated.



Mathematical representation of square waves.

Where the waveform represented by such a series represents a real quantity, such as voltage or current, the components specified by the series are not mathematical fiction. The component frequencies of voltage or current are actually present and can be identified and measured. *See* WAVEFORM, NONSINUSOIDAL.

Application. Square waves are used directly for measurement of the transient response characteristic and indirectly for the frequency response characteristic of electric systems. Mathematically, there is a *unique relationship between the square-wave response* of an electric network and its frequency characteristic.

In addition to the testing of electric circuits and systems, generators of square waves are used as component parts in many electronic systems, including digital computers.

Generation. A square wave may be generated from a sine wave source by a combination of amplification and limiting (*see* CLIPPING CIRCUIT; LIMITER CIRCUIT) whereby only a portion of the wave parallel to the time axis is selected and amplified. More often, square waves are generated by various forms of relaxation oscillators, such as the multivibrator (*see* MULTIVIBRATOR). This is usually the preferred method where short transition times are required. *See* WAVE-SHAPING CIRCUITS. [G.M.C.]

Squash

A member of the plant order Campanulales, squash has been given two distinct definitions. Some taxonomists have limited the term squash to all types and varieties of *Cucurbita maxima*. According to their classification, the varieties popularly referred to as summer squashes are really pumpkins. Other taxonomists, however, define two types of squash: (1) summer squash, which is the edible fruit of any species of *Cucurbita*, commonly *C. pepo*, utilized when immature as a table vegetable, and (2) winter squash, which is the edible fruit of any species of *Cucurbita* utilized when ripe as a table vegetable, in pies, or as feed for livestock, the flesh usually being fine-grained and of mild flavor. These latter definitions are most widely accepted.

Characteristics. Squash and pumpkins are warm-season annual cucurbits native to America (*see* ANNUAL PLANTS). Popular varieties of summer squash (*Cucurbita pepo*) are Yellow Straightneck, Cocozelle, Scallop, and Zucchini. Popular winter squash varieties are Table Queen (*C. pepo*), Butternut (*C. moschata*), and Blue Hubbard, Boston Marrow, and Buttercup (*C. maxima*). Hybrid summer-squash varieties are becoming increasingly popular.

Plants of most summer-squash varieties have a bush habit; most winter-squash varieties have a vining habit. Flowers are monoecious, with the

staminate (male) flowers formed earlier and in greater number *See* FLOWER (BOTANY). Squash and pumpkins will not cross-pollinate with watermelons and muskmelons (*see* REPRODUCTION, PLANT) Although most summer squash will cross with pumpkins and winter squash, and many winter squash will cross with each other, the effect in the year they cross is only on the seed and not the fruit. *See* BREEDING (PLANT); FRUIT (BOTANY).

Propagation is by seed with plantings made in warm soils after frost danger is past. *See* SEED (BOTANY) Occasionally seeds are planted in greenhouses and transplanted to the field when 3–4 weeks old Field spacing varies with plant habit; vining varieties are often in rows that are 8–12 ft apart

Harvesting and storage. Harvesting of summer squash begins before the fruit rinds harden, usually 2–8 days after blossoming and 50–60 days after planting Winter squash are harvested when mature ordinarily 80–120 days after planting

Postharvest temperatures of 50–55°F and low humidity favor prolonged storage. Varieties differ in their keeping quality. Spraying to control foliage in the field reduces storage disease problems Storage favors the conversion of starch to sugar but results in a loss of total carbohydrates and proteins Florida, California, and Texas are important pro-

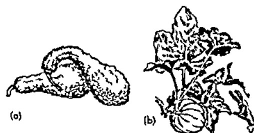


Fig 1 Two kinds of squash, *Cucurbita pepo*. (a) Summer Crookneck. (b) Summer Bergen *(From L H Bailey, The Standard Cyclopedia of Horticulture, vol 3, Macmillan, 1935)*



Fig 2 Mottle and distortion of squash leaves due to mosaic virus.



Fig. 3. (a,b) Two virus-mottled crookneck squash. (c) Healthy fruit.

ducing states. *See* CAMPANULALES; PUMPKIN; VEGETABLE GROWING [H.J C.]

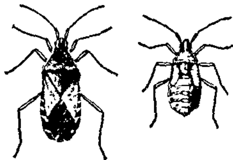
Squash diseases. Diseases cause about a $250,000 loss in the 50,000 acres of squash annually grown in California, Texas, Florida, the midwest, and the mid Atlantic states. The value of the squash crop in these areas exceeds $1,000,000 Seed decay occurs when soils are wet and below 75°F. Seedling growth is good when seed are shallow-sown with adequate soil moisture and are protected with either 6 oz of chloranil or 4 oz of thiram per 100 lb of seed Plants infected with the fungus *Fusarium solani* are stunted, yellowed, wilted, and produce few fruits The tops sometimes break off because the fungus weakens

planting. Seed should not be sown in infested soil, as the fungus lives in the soil for years and attacks both squash and melons. Mosaic viruses, which are both seed-borne and insect-transmitted, mottle the leaves (Fig 2) and distort the fruits (Fig. 3). Virus-free seed, clean weed-free culture, and insect control assure satisfactory yields. *See* BACTERIA; FUNGI, FUNGISTAT AND FUNGICIDE; INSECTICIDE; PLANT VIRUS. [J.T.M]

Bibliography: *See* AGRICULTURAL SCIENCE (PLANT), PLANT DISEASE.

Squash bug

Anasa tristis, an insect of the family Coreidae, order Hemiptera. This is one of the best-known garden pests It punctures the leaves and stems of the squash and sucks the plant juices, leaving the leaves to wilt and die. These bugs also feed on pumpkin and other cucurbits. The adult squash bug appears dark brown, being finely marked with black over a yellow background The eggs are laid in clusters of about 25 on the under sides of leaves Each female will lay several hundred eggs. They hatch into green and black nymphs but in all the other nymphal stages, they are gray Metamorpho-



The squash bug, *Anasa tristis*; length about ¾ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

sis is gradual and from 45 to 60 days are required to complete the life cycle. The species overwinters as the adult. There are six other species of the genus *Anasa* in North America, any of which may be called squash bugs. *See* HEMIPTERA. [J D.B.]

Squid

Any of several species of the suborder Decapoda, order Dibranchia, class Cephalopoda, phylum Mollusca. There are several living species divided into two families. This suborder includes not only the squid, but also the cuttlefish and the small Spirula. These animals all have 10 arms, in contrast to the 8 of the Octopoda.

Best known of the squid is the common squid, *Loligo pealei*, of the Atlantic Coast A similar form on the West Coast is *L. opalescens*.

Squid are worldwide in their distribution. They are used for fish bait, and are important food animals in many parts of the world. When used for food they are usually dried They are also used for fertilizer. Squid are eaten by many marine animals, notably the sperm whale and various fishes

Loligo pealei may be considered typical It is an elongated, cylindrical animal, with a pair of broad fins on the dorsal or posterior end It has 10 arms each of which bears cuplike suckers Two of the arms are modified into long, retractile tentacles. The arms and siphon represent the molluscan foot A mantle surrounds all the internal organs The shell is represented by a slender rod, called the pen, which lies in the anterior part of the animal. The large head bears two conspicuous eyes, superficially identical to the vertebrate eye in structure The head and body join by a neck. Just below the neck is the siphon. The free edge of the mantle forms a loose collar around the neck. In each side of the mantle cavity there is a gill. Ordinarily water passes freely in and out of the mantle cavity by expansion and contraction of the mantle For "jet propulsion," the collar is closed and water is forcibly ejected from the siphon.

The internal anatomy of the squid is otherwise similar to that of a typical clam except that in addition to a liver there is a separate pancreas Two pairs of salivary glands are also present. There is an ink sac opening into the intestine just before the latter terminates as the anus inside the siphon Ink from this sac can be discharged into the water as a means of escaping from an enemy.

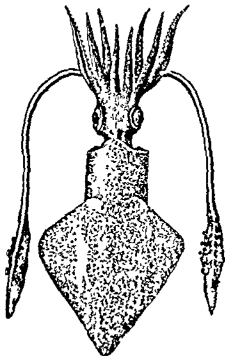
Sexes are separate. The eggs develop in oblong gelatinous capsules from which an animal with the form of the adult hatches.

Squids have a decided ability to change color rather quickly and through a wide range of shades to suit their environment. They eat crustaceans, fish, and other mollusks, which are seized and held firm by the suction cups on the arms and tentacles. Squids frequently occur in great schools, and are among the most common animals of their size in the sea.

The sea arrow, *Ommasthrephes sagittus*, belongs to the family of flying squids and giant squids. They can leap 15 ft or more from the water, and move with great speed through the ocean. They attain a length of 18 in. and prey heavily on schools of herring and mackerel. In turn they are the favorite food of the cod.

The giant squid, *Architeuthis harveyi*, is the largest of all invertebrates. The body has been variously reported from 10 to 15 ft long. The sessile arms are 10 ft long and the tentacles as long as 40 ft. Only a few of these huge animals have been caught, but disgorged arms from captured sperm whales indicate that they may be relatively common in the depths of the ocean, especially in northern waters. There are unverified reports of giant squids much larger than have been supported by actual specimens. This animal is probably the basis for most sea-serpent stories

The cuttle or cuttlefish, *Sepia officinalis*, is related to the squids Its ink is the original India ink It is also the source of the color sepia of painters Its internal shell is the cuttlebone used in bird

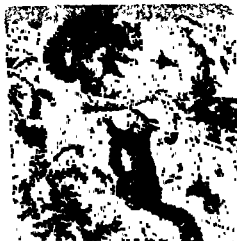


The squid, *Loligo pealei*, length to 8 in (From P Martin Duncan, ed., *Cassell's Natural History*, Cassell)

cages. They are worldwide in their distribution and *quite common* along the Atlantic Coast. They are also used for bait and food. *See* CEPHALOPODA.
[J.D.B.]

Squirrel

Any of many moderate-sized arboreal rodents of the family Sciuridae, found on all the continents except Australia. Squirrels are characterized by long bushy tails and great agility. They are well known for their habit of storing nuts, although their diet is quite varied during the summer, including insects and other small animals as well as a wide variety of plant material. Several species exhibit a fondness for birds and their eggs. The most common of the United States species is the fox squirrel, *Sciurus niger*, and the gray squirrel,



The gray squirrel, Sciurus carolinensis, length 18 in (From P. Martin Duncan, ed., Cassell's Natural History, Cassell)

S. carolinensis. The little red squirrels, of the genus *Tamiasciurus*, are frequently called chickarees. The flying squirrels, *Glaucomys*, are small, nocturnal animals with flattened tails and webs of skin connecting the legs. [illegible]
a [illegible]
b [illegible]
g [illegible]
[J.D.B.]

Stabilization

Any process that minimizes or prevents fluctuations in a quantity or condition. In the navigation of ships and aircraft, a gyroscope provides the [illegible]
[illegible]
o [illegible]
s[illegible]
a [illegible] more powerful stabilization equipment keeps a radar antenna aimed in a desired direction in space despite the roll, pitch, and turning maneuvers of a ship or aircraft.

Stabilization is introduced into vacuum-tube or transistor amplifier stages by means of feedback to reduce distortion by making the amplification substantially independent of electrode voltages and tube constants. With magnetic materials, aging and other treatments produce stabilization of magnetic characteristics.
[J.M.R.]

Stabilizer

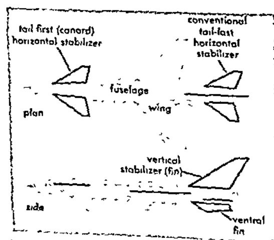
The horizontal or vertical aerodynamic wing surfaces that provide aircraft stability and longitudinal balance in flight. Horizontal and vertical stabilizers (fins) are similar to the aircraft's wing in structural design and function of providing lift at angle of attack to the wind. However, stabilizers are not required to supply lift to overcome aircraft weight during flight, and when the wing-fuselage center of pressure is behind the aircraft center of gravity, the aerodynamic load on the horizontal stabilizer may be downward.

Stabilizer arrangements. Stabilizers may be swept back like wings for improved high Mach-number characteristics. The horizontal stabilizer may be found in a conventional or tail-last arrangement, a canard or tail first arrangement, or it may be dispensed with to give a tailless arrangement. On the other hand, vertical stabilizers are invariably present at the rear of the aircraft in single or multiple units. Aircraft designed for very high speeds may have auxiliary vertical stabilizers or ventral fins below the wing or fuselage, to avoid the large losses in directional stability that occur with conventional arrangements at positive angles of attack at Mach numbers above about 2. Various stabilizer arrangements are illustrated in the accompanying drawing.

Stabilizing function. The stabilizing function of the horizontal and vertical stabilizers may be represented by the equation

$$\frac{\partial C_m}{\partial \alpha} = -\left(\frac{\partial C_L}{\partial \alpha}\right)\frac{S'}{S}\frac{l'}{l}(1-\epsilon_\alpha)\frac{q'}{q}$$

where $\partial C_m/\partial\alpha$ is the dimensionless rate of change of stabilizing or weathervane moment with angle of attack or sideslip, $\partial C_L/\partial\alpha$ is the dimensionless lift-



Various stabilizer arrangements

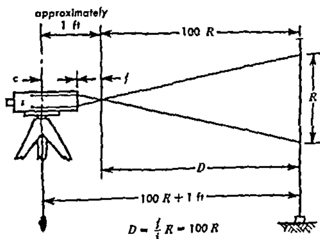
curve slope of the isolated stabilizer based on stabilizer dimensions, and S'/S and l'/l are the ratios of stabilizer area to wing area, and of distance from the stabilizer center of pressure to the aircraft center of gravity to a characteristic length, respectively. The quantities ϵ_α and q'/q are flow conditions at the stabilizer, and are the downwash or sidewash and dynamic pressure, respectively, referred to the free stream.

Control function. On aircraft provided with elevator and rudder control surfaces, the stabilizers may be either fixed to the fuselage or adjustable in incidence at a slow rate (about 1°/sec) for trim. In either case, the stabilizers provide the support for the elevator and rudder, carrying their loads to the fuselage. Supersonic aircraft and guided missiles generally dispense with hinged trailing-edge elevator and rudder control surfaces, using the stabilizers for control as well as stabilization. In this case, the stabilizers are actuated by the primary flight control system at rates up to about 50°/sec. *See* FLIGHT CHARACTERISTICS; FLIGHT CONTROLS. [M J.AB.]

Bibliography: B Etkin, *Dynamics of Flight, Stability and Control*, 1959; C. D. Perkins and R. E. Hage, *Airplane Performance, Stability and Control*, 1949.

Stadia

A method of distance surveying in which measurements are obtained from the interval on a graduated rod intercepted by two parallel lines in the telescope of a transit or other surveying instrument The rod is called a stadia rod; the lines, called *stadia hairs*, are equidistant above and below the horizontal cross hair. Their distance apart i usually is one one-hundredth of the telescope's focal length f. Hence, the distance D from the focal point to an object is 100 times the interval observed on the vertically held rod. For horizontal stadia sighting.



Stadia measurement of distance.

the distance between the focal point and the center of the transit is added to the distance obtained by reading the rod as shown in the illustration. This distance is about 1 ft for external focusing instruments; it is negligible in internal focusing instruments. Further corrections are made for inclined sightings, where the slope distance must be reduced to its horizontal component and where the difference in elevation between the transit point and the rod point also can be obtained. The figures and graduations on the stadia rod can be read by the instrument man through the instrument telescope. *See* SURVEYING. [R H DO.]

Stain (microbiological)

Certain colored organic compounds called dyes, used to stain tissues, cells, cell components, or cell contents. The dyes may be natural or synthetic. The object stained is called the substrate. The small size and transparency of microorganisms makes them difficult to see even with the aid of a high-powered microscope. Stains are one of the methods used to facilitate the examination of these organisms.

Stain classification. Stains may be classified according to their molecular structure. For example, there are the triphenylmethane dyes, like the fuchsins and the methyl violets; the oxazine dyes, like Nile blue; and the thiazine dyes, like thionine and methylene blue

The stains may also be classified according to their chemical behavior into acid, basic, neutral, and indifferent. This classification is of more practical value to the biologist. An acid dye, such as Congo red and eosine, is usually the sodium or potassium salt of a dye acid; a basic dye, such as methylene blue and Nile blue, the chloride or sulfate of a dye base, a neutral dye, such as the eosinate of methylene blue, is a complex salt of a dye acid with a dye base, an indifferent dye, such as Sudan III, is one with little chemical activity. At ordinary pH values, cells stain with basic dyes; at low pH values, they stain with acid dyes

Staining procedures. The staining of microorganisms usually begins by making a smear When it is desired to preserve intercellular relations, the smear may be replaced by a microculture The smear or its substitute is usually fixed by heat and treated with a solution of some suitable biological stain as in simple stains or, in more elaborate procedures, it may be chemically fixed, treated with mordants, stained, partially decolorized, and counterstained In certain procedures, parts of a substrate are chemically removed prior to staining

Smear preparation. The material containing the microorganisms is smeared or spread on the surface of a glass slide Liquid material may be spread by means of a freshly flamed and cooled loop of platinum or some suitable alloy Solid material may be smeared directly, or it may be used to make a suspension in water or some liquid medium, a loopful of which is spread on the slide The film produced is allowed to dry in the air and is fixed by passing the slide, film side up, three or four times over the flame of a Bunsen burner Some cytological procedures require chemical fixation After cooling, the film is stained and examined.

Simple stains. These are procedures in which the smear, or the substrate, is stained with a dye solution for a given length of time, washed with water, air-dried, and examined. The most widely used solutions for simple staining are (1) Löffler's methylene blue, a solution of 0.3 g of methylene blue in 30 ml of 95% ethanol, mixed with 100 ml of 0.01% potassium hydroxide; (2) Ziehl-Neelsen's

crystal violet, prepared by mixing a solution of 0.2 g of crystal violet in 20 ml of 95% ethanol with a solution of 0.8 g ammonium oxalate in 80 ml of distilled water

Differential stains. These are staining procedures that bring out color differences between a substrate and its background, or between different parts of the same substrate. This may be accomplished by simple staining with a metachromatic dye; Nile blue, for example, stains the neutral fat droplets red and the rest of the cell blue. Or it may require controlled decolorization or differentiation with acids, alkalis, neutral solvents, or some other agents. Examples of the latter procedures are Gram's stain and acid-fast stain, and the spore stain. *See* ACID-FAST STAIN; GRAM'S STAIN.

Spore stains. These are based on the fact that the spore does not take up dyes readily, but once stained, resists decolorization. The differentiating agent used may be a dilute solution of an organic

... differentiated with a 0.5% solution of citric acid, and counterstained with a 1% solution of methylene blue. Rickettsiae stain red and tissue cells blue.

Negative stain. Another differential stain is the negative stain. In this procedure the cells appear colorless against a colored background. The cells are suspended in a droplet of the negative stain, such as India ink, water-soluble nigrosine, or Congo red in 1% aqueous solutions, on a glass slide. The droplet is spread out into a thin film and allowed to dry in the air. When pathogenic organisms are to be examined, the preparation is dipped in a mixture of 1 ml concentrated hydrochloric acid (HCl) and 100 ml of 95% ethanol. The acid alcohol changes Congo-red preparations to blue, giving better contrast.

Complex staining procedures. Among complex staining procedures are the flagella and capsule stains. Both include the use of a special mordant to increase the density of their substrates and their affinity for the dye. This also increases the thickness of the flagella to a range that is resolved with the light microscope. There are many procedures for demon...

Nuclear stains. Most nuclear stains are based on the staining properties of nucleic acids. In many organisms, particularly among the bacteria, the cytoplasm is rich in ribonucleic acid. As a result, the entire cell, in young and mature cultures, tends to stain like a nucleus. Consequently, most nuclear stains include a step in which the ribonucleic acid is removed; or the bacteria are grown in media deficient in nitrogen supply to hinder the synthesis of ribonucleic acid.

Ribonucleic acid may be removed by treatment with the enzyme, ribonuclease, or a strong mineral acid. This is the basis of the widely used Feulgen reaction and of the HCl-Giemsa stain. In both methods, the fixed smear is placed in normal HCl, usually at 60°C for about 10 min. In the Feulgen procedure, this is followed by placing the smear for several hours in a solution of basic fuchsin that had been decolorized with sodium or potassium metabisulfite. The nuclei appear purple to violet in a colorless cell.

In the HCl-Giemsa procedure, the smear is stained with dilute Giemsa's stain. Stock solutions of this stain contain chiefly the eosinates of methylene azure and of methylene blue dissolved in a mixture of methyl alcohol and glycerol. It is a valuable cytological stain because of its metachromatism; it stains chromatin purple to red and the cytoplasm blue. *See* MICROBIOLOGICAL METHODS.

[C KN.]

Bibliography: H. J. Conn, *Biological Stains*, 6th ed., 1953; E. Gurr, *A Practical Manual of Medical and Biological Staining Techniques*, 1953; G. Knaysi, *Elements of Bacterial Cytology*, 1951; Society of American Bacteriologists, *Manual of Microbiological Methods*, 1957.

Stained glass

A type of glass used in windows, primarily for churches, and also in the making of smaller ornamental objects. The glass is of three types: pot metal, in which the color is inherent throughout the glass; enameled, in which different colored transparent metallic oxides are fused onto plain or tinted glass, and painted, in which transparent pigments are baked or burned on plain glass. The variously colored, decorated, and shaped pieces of glass are then joined together into pictorial or abstract designs with lead strips, and in larger windows, with iron armatures, or frames. Of the three, mosaic windows made of pot metal, which first appeared in churches during the ninth century in Europe, are by far the most attractive and durable; enameled or painted glass is relatively less permanent. In the nineteenth and early twentieth centuries, particularly in the United States, opalescent (Tiffany) glass was widely used, but is now rarely made. After World War II, a revival of the pot-metal technique of making mosaic stained glass windows was strongly supported by the burgeoning of church building, both in the Americas and in Europe. *See* GLASS AND GLASS PRODUCTS.

[C C O.]

Stainless steel

Steels alloyed with sufficient chromium (over about 9%) to resist effectively the corrosion, oxidation, or rusting which ordinary carbon and low-alloy steels naturally undergo in moist atmosphere or in salt or fresh water (see illustration). The initial oxidation of stainless steels is thought to result in the formation of a very thin, sometimes transparent, and tightly adherent skin of chromium oxide which is impervious to oxygen and which thus effectively prevents progressive oxidation of the metal beneath. In this respect, the effect of oxidation differs from that occurring in ordinary steels, on which is formed a loose, permeable scale which holds moisture and through which oxygen readily diffuses to cause progressive attack of the metal beneath

Because of their high alloy content (chromium sometimes amounting to over 25% and nickel higher still), stainless steels are melted exclusively in electric-arc or electric-induction furnaces In the arc furnace, steel scrap forms the bulk of the charge which, after melting and refining, is brought to the desired composition by the addition of alloys. In the induction furnace, a mixture of scrap and alloys, proportioned so as to yield the desired composition, is melted without any refining operation

Both stainless steels and other steels are produced in the form of wrought products through the intermediate step of casting an ingot which is rolled into bars, plate, or sheet in a rolling mill; extruded or pierced and drawn into pipe; or forged to shape under a press or hammer. They are also cast directly into shape to be put into service without hot working.

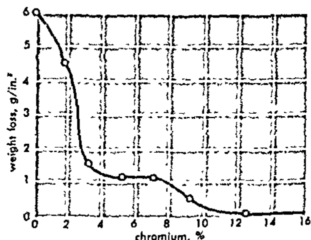
There are two basic types of stainless steel, one depending on chromium alone for oxidation resistance and given numerical designations between 400 and 500 by the American Iron and Steel Institute (AISI) and the other having a large addition of nickel in addition to the chromium and given numerical designations between 300 and 400 by the AISI. The chromium (AISI 400) type is, crystallographically, body-centered cubic, is magnetic, and depending on the relative amounts of carbon and chromium in the alloy, may be hardened, as are other steels, by rapidly cooling it from above a critical temperature which, because of the chromium content, is higher than in ordinary steels When the chromium content is quite high, for example, above 12%, and the carbon content is below about 0.10%, the steel is unhardenable for practical purposes. Such steel becomes hardenable if the carbon is increased, provided the chromium is not also increased. The compositions of typical chromium stainless steels are shown in Table 1.

Chromium stainless steels. These are relatively inexpensive, resistant to ordinary corrosive attack, and capable of being hardened and tempered to produce desirable strength up to 250,000 psi.

Heat-treatment affects the corrosion resistance of the chromium (type 400) stainless steels. Heating the hardened steel to temperatures at which carbon precipitates to form chromium carbides is detrimental to over-all corrosion resistance because the formation of such carbides robs the steel of the protective action of the chromium which they contain. For this reason, a decrease in the carbon content improves the corrosion resistance of steels of a given chromium content. Hardenable grades, containing considerable carbon, have their best resistance to corrosion in the hardened condition. When both hardness and corrosion resistance are sought in the type-400 steels, it is important to select the steel which can be heat-treated to produce the desired hardness with the least sacrifice of corrosion resistance

Table 1. Typical chromium stainless steels

AISI type	% C	% Cr	Characteristics
430	0.12 (max)	14.0–18.0	Nonhardenable
430F	0.12 (max)	14.0–18.0	Se 0.15 min; free machining
440C	0.95–1.20	16.0–18.0	Mo 0.75 max, fully hardenable
446	0.20 (max)	23.0–27.0	Resistant to scaling when hot



Effect of chromium content on loss of weight of low-carbon steel in 4½ years' exposure to New York City atmosphere (From W O Binder and C. M Brown, Proc ASTM, 46,595, 1946)

Chromium-nickel stainless steels. The chromium-nickel (AISI 300) types are nonmagnetic, and are not capable of being hardened and tempered as are the chromium (AISI 400) steels. They have a predominantly face-centered cubic crystalline structure and are commonly referred to as austenitic or "18-8" stainless. The chemical composition of these steels determines whether or not small amounts of magnetic, body centered cubic ferrite may be present The presence of the magnetic phase results from lowering the nickel content, from increasing the chromium content without a commensurate increase in nickel, or from the presence of effective ferrite-forming elements such as silicon, molybdenum, titanium, aluminum, or

Table 2. Typical chromium-nickel stainless steels

AISI type	% Carbon max	% Cr	% Ni	Characteristics
301	0 15	16–18	6 0–8 0	Ferritic, work-hardens
304	0 08	18–20	8 0–12 0	Austenitic, superior corrosion resistance
304L	0 03	18–20	8 0–12 0	Resists sensitization
316	0 08	16–18	10 0–14 0	Mo 2–3%, superior corrosion resistance
316L	0 03	16–18	10 0–14 0	Mo 2–3%, resists sensitization
321	0 08	17–19	9–12	Ti 5 × C%, resists sensitization
347	0 08	17–19	9–13	Nb 10 × C%, resists sensitization
310	0 08	24–26	19–22	High-temperature scale resistant

columbium (niobium). Typical chromium-nickel stainless steel compositions are shown in Table 2.

The resistance to corrosion of the chromium-nickel (type 300) steels is superior to that of the plain chromium (type 400) steels and is best after the steel has been heated to a temperature in excess of about 1800°F, at which all the carbon is in solid solution, and cooled rapidly to prevent its precipitation. In the austenitic steels, the carbides formed in slow cooling, or upon heating the rapidly cooled metal to 1000–1500°F, occupy the crystal boundaries, which become depleted in chromium and vulnerable to attack which can be serious enough to embrittle the metal completely and destroy its metallic continuity When carbides are thus precipitated in the crystal boundaries, the steel is said to be sensitized There are two practical preventive measures to overcome this condition. One of these is to alloy the metal, during melting, with titanium [illegible]

[illegible] with the chromium The other method is to remove carbon as completely as possible during melting in the electric-arc furnace. Steels so made have carbon below 0 03% and are given the terminal designation "L" as in types 304L and 316L

Although unhardenable, in the usual sense of the word, certain of the [illegible] notably typ[illegible] tensile stre[illegible] strengths a[illegible] containing [illegible] ferrite, are cold worked [illegible] zero [illegible] [illegible] absence of ferrite. See ALLOY; CORROSION; STEEL

[F.B.F.]

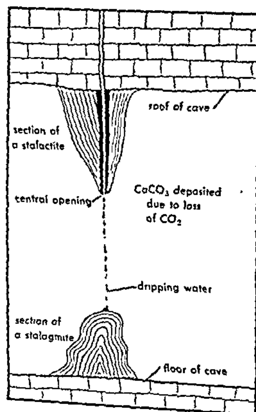
Stalactites and stalagmites

Stalactites, stalagmites, dripstone, and flowstone are travertine deposits in limestone caverns, formed by the evaporation of waters bearing calcium carbonate. Stalactites grow down from the roofs of caves and tend to be long and thin, with hollow cores. The water moves down the core and precipitates at the bottom, slowly extending the length while keeping the core open for more water to move down. Stalactites are banded concentrically to the center.

Stalagmites grow from the floor up and are commonly found beneath stalactites; they are formed from the evaporation of the same drip of water that forms the stalactite. Stalagmites are thicker and shorter than stalactites and have no central hollow core The banding of stalagmites is parallel to the surface at the time of deposition of each band.

Dripstone and flowstone are travertine accumulations of much the same origin as stalactites and stalagmites (evaporation from cave waters) but are irregular. Their irregularity is due to deposition on irregular sloping walls or surfaces in the cave rather than drips from the roof. See CALCITE; CAVE; LIMESTONE; TRAVERTINE. [R.S.]

Bibliography: V C. Allison, The growth of stalagmites and stalactites, *J. Geol.*, 31(2):106–125, 1923



Stalactite and stalagmite deposits Black areas along central opening represent deposits made by water as it dripped from the crack in limestone Outside layers were deposited by water flowing over surface. Drops of water may fall to the floor and create a stalagmite. (After E B. Branson, W A Tarr, C C. Branson, and W. D. Keller, *Introduction to Geology*, McGraw Hill, 3d ed., 1952)

Stall-warning indicator

An instrument which indicates the highest angle of attack, or the lowest indicated air speed, which may be safely flown. The warning signal, either visual or—preferably—auditory, may be obtained from an angle-of-attack indicator (*see* YAW INDICATOR). Alternatively, either a pressure-sensing element or a vane is installed on the leading or trailing edge of the wing. Pressure taps, connected to the pressure-sensing elements, give warning when pressure changes indicate the onset of stall. The vanes are arranged to flip over as the air circulation over the wing approaches stall conditions. The vanes make an electric contact to operate a warning device. [W.G.B.]

Standard

An accepted reference sample used for establishing a unit for the measurement of physical quantities. A physical quantity is specified by a numerical factor and a unit; for example, a mass might be expressed as 8 g, a length as 6 cm. Here the gram is a mass unit defined in terms of the international prototype kilogram, which serves as the primary standard of mass; similarly, the centimeter is defined in terms of the international prototype meter, which provides the primary standard of length

The National Bureau of Standards in the United States and comparable laboratories in other countries are responsible for maintaining accurate secondary standards for various physical quantities. *See* ELECTRICAL STANDARDS; KILOGRAM; METER (UNIT), PHYSICAL MEASUREMENT, TIME, WAVELENGTH STANDARDS. [D.W.I.]

Standing wave

A wave which is formed when the boundary conditions on traveling waves are such that the transmitted and reflected waves add up so that the amplitudes are constant at a fixed position in space but vary from point to point. The points in space where the amplitudes are zero are called nodes. Antinodes are points where amplitude is a maximum.

For the case of sinusoidal waves, the standing waves assume a particularly simple form. Consider two waves q_1 and q_2 of equal amplitude A

$$q_1 = A\cos\left(2\pi\nu t - \frac{2\pi x}{\lambda} + \phi_1\right)$$

$$q_2 = A\cos\left(2\pi\nu t - \frac{2\pi x}{\lambda} + \phi_2\right)$$

where ν is the frequency, t is the time, x is the direction of propagation, λ is the wavelength, and ϕ is the phase angle. The sum of these two waves is

$$q_1 + q_2 = A\left[\cos\left(2\pi\nu t - \frac{2\pi x}{\lambda} + \phi_1\right) + \cos\left(2\pi\nu t - \frac{2\pi x}{\lambda} + \phi_2\right)\right]$$

If the boundary conditions are such that $\phi_1 = \phi_2$ and if $\phi = 0$,

$$q_1 + q_2 = 2A\cos\left(2\pi\nu t - \frac{2\pi x}{\lambda}\right)$$

$$= 2A\cos\frac{2\pi x}{\lambda}\cos 2\pi\nu t$$

Thus amplitude at any position is $2A\cos(2\pi x/\lambda)$, and for positions such that $2\pi x/\lambda = (n + \frac{1}{2})\pi$, where n is an integer, the amplitude is zero; for $2\pi x/\lambda = n\pi$ the amplitude is a maximum

Acoustic waves can form standing waves in closed systems with the proper dimensions and boundary conditions The ratio of the maximum to the minimum effective sound pressure, called the standing-wave ratio, can be used to measure the acoustic impedance of the system. Electromagnetic waves can form standing waves in wave guides and transmission lines, and the standing-wave ratio can be used to determine the impedance of the wave guide or transmission line.

Standing waves are a special case of stationary waves in which at least one of the enclosure terminations absorbs a part of the energy of the incident waves, as well as reflecting a portion, resulting in a net power loss from the source. *See* STANDING-WAVE DETECTOR; STATIONARY WAVE; TRANSMISSION LINES; WAVE (PHYSICS), WAVE EQUATION, WAVE GUIDE; WAVE MOTION. [W.J.G.]

Standing-wave detector

An electric indicating instrument used for detecting standing waves along a transmission line or in a wave guide and measuring the resulting standing-wave ratio. It can also be used to measure the wavelength and hence the frequency of an electromagnetic wave in a line. The detecting device is usually a bolometer, thermocouple, or crystal, connected to an indicating meter directly or through an amplifier The detecting device is moved along the line while observing the meter indication; the positions along the line at which maximum and minimum readings are obtained correspond to the nodes and antinodes of the standing wave that is produced by transmitted and reflected waves of equal frequency moving in opposite directions The reflected wave is generated at a discontinuity in the transmission line or wave guide *See* WAVELENGTH MEASUREMENT; WAVEMETER [J.MR.]

Staphylococcus

A genus of the bacterial family Micrococcaceae of the order Eubacteriales The staphylococci are parasites of man, their usual habitat being the nasopharynx and skin They may cause boils, osteomyelitis, wound infections, and some cases of septicemia, pneumonia, and kidney infections They are gram-positive, spherical, pathogenic bacteria, about 1 micron in diameter They typically occur in grapelike clusters. Staphylococci grow readily on the usual laboratory culture media; meat infusion agar, containing human or animal blood, is frequently employed for their isolation Staphylococci are relatively resist [illegible] adverse

physical and chemical agents. Since 1944, following the introduction of antibiotics in the treatment of infectious diseases, antibiotic-resistant strains of staphylococci have become increasingly numerous Infections due to these resistant strains have raised serious problems of treatment and control

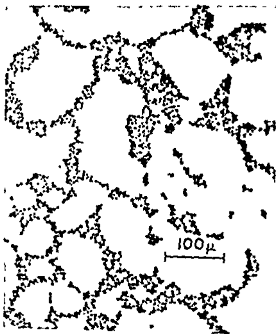
Staphylococci produce pigments which range from orange or golden-yellow to pale cream or white. This has served to identify and classify them. Several species were long recognized and designated as *Staphylococcus aureus*, golden; *S. albus*, white; and *S citreus*, lemon yellow. Between 1948 and 1957 staphylococci were designated by many bacteriologists as the species *Micrococcus pyogenes*, with subspecies being identified by the pigment produced, as in the name *M pyogenes* var. *aureus* In 1957 the term *S. aureus* was reintroduced as the specific term for pathogenic staphylococci, with the understanding that the degree of pigmentation may be variable. The characteristic pigmentation of cultures, the microscopic appearance of the cocci and the ability of a culture to cause clotting of blood plasma, serve to identify pathogenic staphylococci.

Toxins and enzymes. Staphylococci produce several serologically distinct toxins and enzymes which are assumed to contribute to their capacity to produce disease Confirmation of the precise role of these substances in the pathogenesis of staphylococcal infection awaits further studies with highly purified preparations Among these substances, exotoxin produces intense destruction of the local tissues when injected into the skin of experimental animals and is rapidly fatal when injected [illegible]

he [illegible]

s[illegible] Each is distinguished by its capacity to hemolyze the red blood cells of certain animal species This is usually demonstrated by the development of clear areas surrounding colonies on nutrient agar The medium contains the blood of a susceptible species, and the clearing represents partial to complete hemolysis The α-hemolysin is produced by many strains that are pathogenic for man, the β-hemolysin is produced chiefly by strains from animal sources and only occasionally by strains from human sources The δ hemolysin is formed by strains that produce either α- or β-hemolysin The relation of γ-hemolysin to the others remains to be determined The significance of hemolysin both as a criterion of pathogenicity and as a factor in the production of disease, is in dispute

Enterotoxin This is distinct from exotoxin and other staphylococcal products, and is responsible for the acute gastrointestinal symptoms of staphylococcal food poisoning Not all pathogenic staphylococci produce this toxin However, enterotoxigenic staphylococci are widely distributed, and staphylococcal food poisoning is probably the most common of all forms of bacterial food poisoning. The toxin is produced in the contaminated food



Morphology and grouping of staphylococci from an 18-hour culture, gram-stained preparation. (University of Virginia Hospital)

during rapid growth of the staphylococci and is already preformed when the food is eaten. The responsible staphylococci are often derived from the nose or skin, or from an infection on the skin, of a food handler. In some cases of food poisoning due to milk or milk products, the offending cocci are derived from staphylococcal mastitis in a cow or goat supplying the milk

Leukocidin. This is a toxin which destroys the [illegible]

Coagulase The production of this enzyme is confined essentially to pathogenic staphylococci; it is not formed by nonpathogenic strains. In the test tube, coagulase produces a clot in the blood plasma of some animal species, acting in combination with a coagulase-reacting factor in the blood. A simple coagulase test, using human or rabbit plasma, is at present the most reliable in vitro test for determination of the potential pathogenicity of staphylococci. It is generally held that coagulase acts in some way to enable the cocci to become established in the body tissues, following which other toxic or enzymic factors come into play. The exact manner in which it acts remains to be demonstrated Significantly, only those animal species (including man) whose blood contains the coagulase-reacting factor are susceptible to staphylococcal infection.

Hyaluronidase This enzyme is produced by most pathogenic staphylococci It breaks down certain constituents of the intracellular ground substance of many body tissues. Hypotheses concerning the role of the enzyme include neutralization of substances (hyaluronic acid) in the tissues which would normally restrict the action of staphy-

Stall-warning indicator

An instrument which indicates the highest angle of attack, or the lowest indicated air speed, which may be safely flown. The warning signal, either visual or—preferably—auditory, may be obtained from an angle-of-attack indicator (*see* YAW INDICATOR). Alternatively, either a pressure-sensing element or a vane is installed on the leading or trailing edge of the wing. Pressure taps, connected to the pressure-sensing elements, give warning when pressure changes indicate the onset of stall. The vanes are arranged to flip over as the air circulation over the wing approaches stall conditions. The vanes make an electric contact to operate a warning device [W.C.B.]

Standard

An accepted reference sample used for establishing a unit for the measurement of physical quantities. A physical quantity is specified by a numerical factor and a unit; for example, a mass might be expressed as 8 g, a length as 6 cm. Here the gram is a mass unit defined in terms of the international prototype kilogram, which serves as the primary standard of mass; similarly, the centimeter is defined in terms of the international prototype meter, which provides the primary standard of length

The National Bureau of Standards in the United States and comparable laboratories in other countries are responsible for maintaining accurate secondary standards for various physical quantities. *See* ELECTRICAL STANDARDS; KILOGRAM; METER (UNIT); PHYSICAL MEASUREMENT, TIME; WAVELENGTH STANDARDS. [D.W.T.]

Standing wave

A wave which is formed when the boundary conditions on traveling waves are such that the transmitted and reflected waves add up so that the amplitudes are constant at a fixed position in space but vary from point to point. The points in space where the amplitudes are zero are called nodes Antinodes are points where amplitude is a maximum

For the case of sinusoidal waves the standing waves assume a particularly simple form Consider two waves q_1 and q_2 of equal amplitude A

$$q_1 = A\cos\left(2\pi\nu t - \frac{2\pi x}{\lambda} + \phi_1\right)$$
$$q_2 = A\cos\left(2\pi\nu t - \frac{2\pi x}{\lambda} + \phi_2\right)$$

where ν is the frequency, t is the time, x is the direction of propagation, λ is the wavelength, and ϕ is the phase angle. The sum of these two waves is

$$q_1 + q_2 = A\left[\cos\left(2\pi\nu t - \frac{2\pi x}{\lambda} + \phi_1\right) + \cos\left(2\pi\nu t - \frac{2\pi x}{\lambda} + \phi_2\right)\right]$$

If the boundary conditions are such that $\phi_1 = \phi_2$ and if $\phi = 0$,

$$q_1 + q_2 = 2A\cos\left(2\pi\nu t - \frac{2\pi x}{\lambda}\right)$$
$$= 2A\cos\frac{2\pi x}{\lambda}\cos 2\pi\nu t$$

Thus amplitude at any position is $2A\cos(2\pi x/\lambda)$, and for positions such that $2\pi x/\lambda = (n + \frac{1}{2})\pi$, where n is an integer, the amplitude is zero; for $2\pi x/\lambda = n\pi$ the amplitude is a maximum.

Acoustic waves can form standing waves in closed systems with the proper dimensions and boundary conditions. The ratio of the maximum to the minimum effective sound pressure, called the standing-wave ratio, can be used to measure the acoustic impedance of the system Electromagnetic waves can form standing waves in wave guides and transmission lines, and the standing-wave ratio can be used to determine the impedance of the wave guide or transmission line

Standing waves are a special case of stationary waves in which at least one of the enclosure terminations absorbs a part of the energy of the incident waves, as well as reflecting a portion, resulting in a net power loss from the source. *See* STANDING-WAVE DETECTOR; STATIONARY WAVE, TRANSMISSION LINES; WAVE (PHYSICS), WAVE EQUATION, WAVE GUIDE; WAVE MOTION [W.J.G.]

Standing-wave detector

An electric indicating instrument used for detecting standing waves along a transmission line or in a wave guide and measuring the resulting standing-wave ratio It can also be used to measure the wavelength and hence the frequency of an electromagnetic wave in a line. The detecting device is usually a bolometer, thermocouple, or crystal, connected to an indicating meter directly or through an amplifier The detecting device is moved along the line while observing the meter indication; the positions along the line at which maximum and minimum readings are obtained correspond to the nodes and antinodes of the standing wave that is produced by transmitted and reflected waves of equal frequency moving in opposite directions The reflected wave is generated at a discontinuity in the transmission line or wave guide *See* WAVELENGTH MEASUREMENT, WAVEMETER [J MR]

Staphylococcus

A genus of the bacterial family Micrococcaceae of the order Eubacteriales. The staphylococci are parasites of man, their usual habitat being the nasopharynx and skin. They may cause boils, osteomyelitis, wound infections, and some cases of septicemia, pneumonia, and kidney infections. They are gram-positive, spherical, pathogenic bacteria, about 1 micron in diameter They typically occur in grapelike clusters. Staphylococci grow readily on the usual laboratory culture media, meat-infusion agar, containing human or animal blood, is frequently employed for their isolation. Staphylococci are relatively resistant to adverse

Table 1. *The 26 nearest stars (from P. van de Kamp)*[a]

Name	Parallax, seconds of arc	Distance, light years	Annual proper motion, seconds of arc	Radial velocity, km/sec	Transverse velocity, km/sec	Apparent magnitude and spectrum	Absolute magnitude
Sun						−26.8	+4.7
α Centauri[b]	0.760	4.3	3.68	−25	23	+ 0.3 G2 (1.7 K5)	4.7(6.1)[c]
Barnard's star	.545	6.0	10.30	−108	90	9.5 M5	13.2
Wolf 359	.421	7.7	4.81	+13	51	13.5 M6e	16.6
Luyten 726–8	.410	7.9	3.35	+29	38	12.5 M6e (13.0 M6e)	15.6(16.1)
Lalande 21185	.398	8.2	4.78	−86	57	7.5 M2	10.5
Sirius[c]	.375	8.7	1.32	−8	16	− 1.6 A0 (7.1 DA)	1.3(10.0)
Ross 154	.351	9.3	0.67	−4	9	10.6 M5e	13.3
Ross 248	.316	10.3	1.58	−81	23	12.2 M6e	14.7
ε Eri	.303	10.8	0.97	+15	15	3.8 K2	6.2
Ross 128	.298	10.9	1.40	−13	22	11.1 M5	13.5
61 Cyg	.293	11.1	5.22	−64	84	5.6 K6 (6.3 M0)	7.9(8.6)
Luyten 789–6	.292	11.2	3.27	−60	53	12.2 M6	14.5
Procyon	.288	11.3	1.25	−3	20	0.5 F5 (10.8 DA?)	2.8(13.1)
ε Indi	.285	11.4	4.67	−40	77	4.7 K5	7.0
Σ 2398	.280	11.6	2.29	+1	38	8.9 M4 (9.7 M4)	11.1(11.9)
Groombr 34	.278	11.7	2.91	+14	19	8.1 M2e (10.9 M4e)	10.3(13.1)
τ Cet	.275	11.8	1.92	−16	33	3.6 G4	5.8
Lacaille 9352	.273	11.9	6.87	+10	118	7.2 M2	9.4
+5° 1668	.263	12.4	3.73	+26	67	10.1 M4	12.2
Lacaille 8760	.255	12.8	3.46	+23	61	6.6 M1	8.6
Kapetyn's star	.251	13.0	8.79	+242	166	9.2 M0	11.2
Ross 614	.251	13.1	0.97	+24	18	11.3 M5e (14.8)	13.3(16.8)
Kruger 60	.249	13.1	0.87	−24	16	9.9 M4 (11.4 M5e)	11.9(13.4)
−12° 4523	.244	13.4	1.24	−13	21	10.0 M5	11.9
vMa 2[d]	0.236	13.8	2.98	+70	59	12.3 DG	14.2

[a] In 26 listings there are 36 *individual* stars, and some may have still undiscovered faint companions.
[b] α Centauri has a second companion (11 M5e) of absolute magnitude +15.4
[c] [illegible]
[d] [illegible]

Near stars. The nearest stars have been so identified by their large angular proper motions in the plane of the sky and by subsequent measurements of parallax *See* Parallax (astronomy) The 26 nearest stars are listed in Table 1 together with relevant data A few more, intrinsically faint stars may exist at or within this limit of distance Table 1 also gives the transverse motions, in kilometers per second, derived from the proper motion and distance of the star, and the radial velocity, which is from the measured Doppler shift. Many stars have a total space motion of over 100 km/sec with respect to the Sun Such objects are called high velocity stars and belong, according to the nomenclature of Walter Baade, to population II, that is stars found in the central region of a galaxy The slower-moving stars are younger, and may be members of population I like the Sun, such stars are found in the outer spiral region of a galaxy The absolute visual magnitudes show that 36 of the 40 nearby stars are intrinsically fainter than the Sun. In addition, about half are in multiple systems, doubles or triples The stars have spectral types which put them on the main sequence or dwarf branch No red giant or supergiant is included although there are three white dwarfs in our immediate neighborhood Figure 1, a Hertzsprung-Rus-

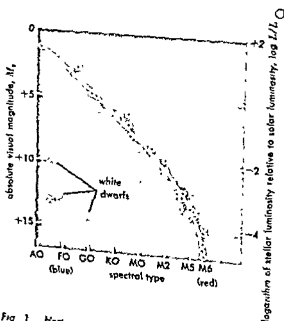
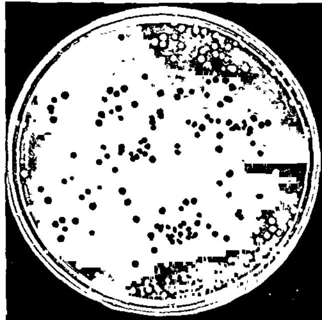


Fig 1 Hertzsprung Russell diagram for nearby stars shows main sequence and white dwarfs.



Colonies of staphylococci growing on nutrient agar.

lococcal toxins, and the enhancement of mixed infections by staphylococci and viruses.

Staphylococcal infections. Staphylococci are pyogenic, or pus-forming, bacteria. Typically they tend to produce circumscribed lesions, in the form of abscesses, which often occur in the skin and immediately underlying tissues, but which may be found in nearly any tissue or organ of the body Staphylococci are the cause of furuncles, or boils, and of carbuncles, and are responsible for about 95% of all cases of the bone infection, osteomyelitis, Most of the common infections of wounds, cuts, or burns are due to staphylococci. They may also produce some cases of septicemia, meningitis, pneumonia, kidney infection, and an assortment of other conditions. Since about 1950, outbreaks of purulent skin infections largely due to antibiotic-resistant staphylococci, have occurred in newborn babies, and have produced breast abscesses in nursing mothers. These outbreaks have occurred in various parts of the world Staphylococcal pneumonia is a frequent complication of epidemic influenza. The responsibility of staphylococci for many outbreaks of food poisoning was mentioned above. Enterotoxigenic staphylococci have been incriminated in some cases of enteritis following therapeutic use of antibiotics. Staphylococci also may produce infections of some domestic animals and birds.

Investigations of staphylococcal infections are aided by a method of identification and differentiation of cultures known as bacteriophage typing. *Staphylococci, like other bacteria,* such as *Escherichia coli,* are susceptible to infection by a virus, or bacteriophage. The bacteriophages are specific, not only for the organism, but also for strains of the organism. A determination is made of the susceptibility of cultures to a series of staphylococcal bacteriophages. Identical or closely similar cultures exhibit identical or closely similar susceptibility to the bacteriophages. Unrelated cultures show distinct differences in their patterns of susceptibility. Bacteriophage typing is especially useful as a tool in the epidemiological study of staphylococcal infections. It permits identification of an epidemic strain, the demonstration of its presence in carriers, and the determination of its source and routes of dissemination. *See* BACTERIOLOGY, MEDICAL; BACTERIOPHAGE; BLOOD PLATE HEMOLYSIS; ESCHERICHIA; FOOD POISONING, BACTERIAL; HYALURONIDASE; LEUKOCIDIN. [J.F.B.]

Bibliography: R. J. Dubos (ed.), *Bacterial and Mycotic Infections of Man,* 3d ed., 1958; G. S Wilson and A. A. Miles (eds.), *Topley and Wilson's Principles of Bacteriology and Immunity,* vol. 1, 4th ed., 1955.

Star

A celestial body, consisting of a large, self-luminous mass of hot gas held together by its own gravity The Sun is a typical star; its physical parameters are

$$\text{Radius } R_{\odot} = 6.9 \times 10^{10} \text{ cm}$$
$$\text{Mass } M_{\odot} = 2 \times 10^{33} \text{ g}$$
$$\text{and} \quad \text{Luminosity } L_{\odot} = 4 \times 10^{33} \text{ ergs/sec}$$

Its mean density is 1.45 g/cm³, and its central density is about 150 g/cm³. The surface temperature is 5750°K, the mean temperature about 5×10^{6}°K, and the central temperature 15×10^{6}°K. In spite of the high density, the gas in the Sun is almost completely ionized from surface to center.

Composition and distribution. The composition by weight of the average star is about 70% hydrogen, 28% helium, 1.5% carbon, nitrogen, oxygen, and neon, and 0.5% iron group and heavier elements

The stars contain by far the largest fraction of the mass of the universe; the parameters given above describe their average condition Stars are born, produce nuclear energy, evolve, and eventually die. Their life spans range from 10^{6} years for a star of high luminosity, to 10^{10} for the Sun, and up to 10^{13} for the faintest main-sequence stars The oldest known stars in our galaxy are nearly 10^{10} years old

The nomenclature for the identification of stars gives their general location and brightness The brighter stars were named and the sky subdivided into constellations, Greek alphabet letters generally describe them, with α representing the brightest star, visually, in a given constellation Thus Betelgeuse, or α Orionis, is the brightest star in Orion. Fundamentally a star is defined by its coordinates on the celestial sphere, right ascension and declination, and its brightness, or apparent magnitude (*see* CELESTIAL SPHERE; MAGNITUDE, STELLAR). Because of the precession of the equinoxes, celestial coordinates must be specified for a given epoch About 6000 stars are visible to the naked eye, but over 10^{11} exist in our own galaxy. Useful catalogs give positions, brightnesses, motions, parallaxes, spectral types, velocities, and other properties of the stars.

Table 2. The 25 brightest stars (from H L. Johnson)*

Star	Name	Spectrum	Absolute visual magnitude, M_v	Visual brightness, V	Color index, B − V	Remarks
α CMa	Sirius	A1 V	+1.4	−1.43	0.00	
α Car	Canopus	F0 Ia	−4.5	−0.73	+0.15	
α Cen		G2 V	+4.7	−0.27	+0.66	Double
α Boo	Arcturus	K2 IIIp	−0.1	−0.06	+1.23	
α Lyr	Vega	A0 V	+0.5	+0.04	0.00	
α Aur	Capella	G8 IIIp	−0.6	+0.09	+0.80	Spectroscopic binary, double
β Ori	Rigel	B8 Ia	−7	+0.15	−0.04	Double
α CMi	Procyon	F5 IV-V	+2.7	+0.37	+0.41	
α Eri	Achernar	B3 V	−2	+0.53	−0.16	
β Cen		B0.5 V	−4	+0.66	−0.21	Double
α Ori	Betelgeuse	M2 Iab	−5	+0.7	+1.87	Variable
α Aql	Altair	A7 IV-V	+2.2	+0.80	+0.22	
α Tau	Aldebaran	K5 III	−0.7	+0.85	+1.52	Variable, double
α Cru		B0.5 V	−4	+0.87	−0.24	Double
α Sco	Antares	M1 Ib	−4	+0.98	+1.80	Double, variable
α Vir	Spica	B1 V	−3	+1.00	−0.23	Spectroscopic binary
α PsA	Fomalhaut	A3 V	+1.9	+1.16	+0.09	
β Gem	Pollux	K0 III	+1.0	+1.16	+1.01	
α Cyg	Deneb	A2 Ia	−7	+1.26	+0.09	
β Cru		B0.5 IV	−4	+1.31	−0.23	
α Leo	Regulus	B7 V	−0.7	+1.36	−0.11	Double
ε CMa	Adhara	B2 II	−5	+1.49	−0.17	
α Gem	Castor	A0	+0.9	+1.59	+0.05	Double, spectroscopic binary
λ Sco	Shaula	B2 IV	−3	+1.62	−0.23	
γ Ori	Bellatrix	B2 III	−4	+1.64	−0.23	

* The spectra are from H L. Johnson and Morgan, colors and magnitudes are photoelectric, V being the equivalent of a visual brightness, and B − V being a blue *minus* visual color index The absolute visual magnitudes M_v are based on measured parallaxes, when only one significant figure is given, however, they are only estimates.

tral classification includes a vast majority of the stars and represents a sequence of decreasing reversing layer temperature, from left to the right (Fig 4) There are several side sequences whose temperatures approximately correspond with those of normal stars in which apparent abundance differences exist Both dwarf and giant stars exist over a wide range of spectral type Decimal subdivisions are used, prefixes and suffixes are added in more refined analyses such as e for emission lines, n for broad lines because of rotation, q for novalike, p for peculiar, d for dwarf, D for white dwarf, sg for subgiant, g for giant, and c for supergiant Almost 300 000 stars have been classified on this system

A more refined system of spectral classification has recently been widely used, developed by W W

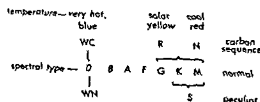


Fig 4 Spectral classification from Henry Draper catalog

Morgan, with P C. Keenan and other collaborators. In this system, employing spectra at 120 A/mm, a two-parameter set of criteria provides by inspection both spectral type and estimated luminosity The luminosity is indicated by a suffix ranging from Ia, extremely bright supergiant, to III, normal giant, to V, main-sequence or dwarf star Thus the designation G2 V represents a star like the Sun, G2 III a giant star of nearly the same temperature, with certain luminosity-sensitive features enhanced

Peculiar spectra exist, such as the carbon stars (old name R and N, new type C) and S stars which have about the same temperature as K or M stars Another very important type of peculiarity has been found associated with stars with magnetic fields (see STELLAR MAGNETIC FIELD).

The percentage of stars of different spectral types in the Henry Draper catalog is roughly as follows: B, 3%, A, 27%; F, 10%; G, 16%; K, 37%; M, 7%; other types are rare This is a selection of stars by apparent brightness and does not represent their true distribution in space, which from the data in Table 1 heavily favors the late-type M dwarfs

Temperature and luminosity. The major part of the differences in appearance of stellar spectra is caused by the change in the surface temperature

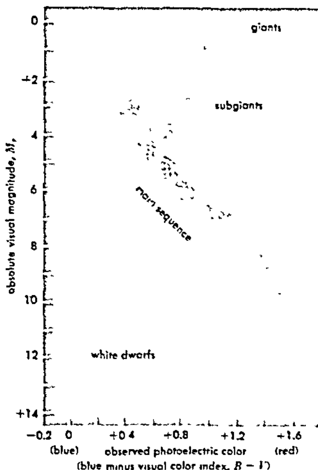


Fig 2 Main-sequence stars, subgiants, and giants.

sell or H-R diagram, is a means for plotting the relation between the luminosity and surface temperature of stars. Along the main sequence through which the line is drawn in the diagram, the stars are distinguished chiefly by mass; luminosity is a steep function of the mass. Nearby stars, for which data are most accurate, do not include all types. To include a greater variety of stars, the less accurate data for stars out to 20 parsecs are plotted (*see* PARSEC). The relation between absolute visual magnitude and observed photoelectric color for the stars closer than 20 parsecs, that is, within approximately 60 light years, shows the main sequence in more detail for the somewhat brighter stars and the red giants and subgiants (Fig. 2). The use of photoelectric color, like that of spectral type, is essentially an arrangement by surface temperature.

Bright stars. The apparently brighter stars are listed in Table 2. Because stars of high intrinsic luminosity can be seen at great distances, Table 2 includes many such stars, including giants and supergiants. For many of these the parallax is too small to be measured directly and the luminosities [illegible] …imate. Similarly, close visual dou- [illegible] the [illegible] …other [illegible] red- [illegible] 3). [illegible] …arge majority of cases shows absorption lines superposed on a continuous background. The interior of the star is at high temperature and pressure; its spectrum resembles that of a black body. The star shades off into space through a reversing layer or stellar atmosphere in which the continuous spectrum and absorption lines are formed. In the Sun this reversing layer is about 400 km thick; its base is at about 5750°K, and its outer layer is near 4200°K. The continuum is formed at a depth equal to the mean free path for an average quantum. The emergent continuum resembles, but is not identical with, a black body. The atoms in the reversing layer produce absorption lines due to the existence of this temperature gradient. Outside the normal reversing layer there may be a temperature inversion in the low-density chromosphere and corona (*see* SUN). The temperature of the corona eventually reaches 1,000,000°K. The emission lines of these outer layers affect only slightly the spectrum of the integrated light of a star.

Stellar spectra are normally obtained with a slit spectrograph by black and white photography. Astronomical spectrographs largely employ plane gratings, operate in the region λ3000–7000 A, and provide dispersions ranging from 1 to about 400 A/mm. Wavelength standards are provided by a comparison spectrum impressed during the exposure of a plate, usually by a laboratory source imaged at both ends of the spectrograph slit. Standards for photometry of the lines are provided by *plate calibration devices*. Because a point source, like a star, gives only a narrow streak of spectrum with the stigmatic spectrographs used, the spectra are suitably widened by allowing the star image to trail up and down the slit. Low-resolution spectra and some high-resolution spectra of bright stars and of the Sun can be obtained by photoelectric scanning at the telescope.

Spectral classification. The spectral classification of a star gives in a simple symbolic form the essential features of its complex spectrum. By inspection, at dispersions ranging from 100 to 400 A/mm, many features vary in a smooth way from one star to another. This variation is correlated with the colors of the stars. As a result spec-

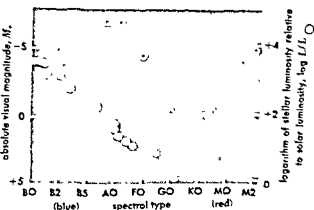


Fig 3. Diagram of brightest stars shows main sequence as in Fig 1 plus branches of other type stars.

Table 2. The 25 brightest stars (from H. L. Johnson)*

Star	Name	Spectrum	Absolute visual magnitude, M_v	Visual brightness, V	Color index, B − V	Remarks
α CMa	Sirius	A1 V	+1.4	−1.43	0.00	
α Car	Canopus	F0 Ia	−4.5	−0.73	+0.15	
α Cen		G2 V	+4.7	−0.27	+0.66	Double
α Boo	Arcturus	K2 IIIp	−0.1	−0.06	+1.23	
α Lyr	Vega	A0 V	+0.5	+0.04	0.00	
α Aur	Capella	G0 IIIp	−0.6	+0.09	+0.80	Spectroscopic binary, double
β Ori	Rigel	B8 Ia	−7	+0.15	−0.04	Double
α CMi	Procyon	F5 IV-V	+2.7	+0.37	+0.41	
α Eri	Achernar	B3 V	−2	+0.53	−0.16	
β Cen		B0.5 V	−4	+0.66	−0.21	Double
α Ori	Betelgeuse	M2 Iab	−5	+0.7	+1.87	Variable
α Aql	Altair	A7 IV-V	+2.2	+0.80	+0.22	
α Tau	Aldebaran	K5 III	−0.7	+0.85	+1.52	Variable, double
α Cru		B0.5 V	−4	+0.87	−0.24	Double
α Sco	Antares	M1 Ib	−4	+0.98	+1.80	Double, variable
α Vir	Spica	B1 V	−3	+1.00	−0.23	Spectroscopic binary
α PsA	Fomalhaut	A3 V	+1.9	+1.16	+0.09	
β Gem	Pollux	K0 III	+1.0	+1.16	+1.01	
α Cyg	Deneb	A2 Ia	−7	+1.26	+0.09	
β Cru		B0.5 IV	−4	+1.31	−0.23	
α Leo	Regulus	B7 V	−0.7	+1.36	−0.11	Double
ε CMa	Adhara	B2 II	−5	+1.49	−0.17	
α Gem	Castor	A0	+0.9	+1.59	+0.05	Double, spectroscopic binary
λ Sco	Shaula	B2 IV	−3	+1.62	−0.23	
γ Ori	Bellatrix	B2 III	−4	+1.64	−0.23	

* The spectra are from H. L. Johnson and Morgan colors and [illegible]

tral classification includes a vast majority of the stars and represents a sequence of decreasing reversing-layer temperature, from left to the right (Fig. 4). There are several side sequences whose temperatures approximately correspond with those of normal stars, in which apparent abundance differences exist. Both dwarf and giant stars exist over a wide range of spectral type. Decimal subdivisions are used, prefixes and suffixes are added in more refined analyses such as e for emission lines, n for broad lines because of rotation, q for novalike, p for peculiar, d for dwarf, D for white dwarf, sg for subgiant, g for giant, and c for supergiant. Almost 300,000 stars have been classified on this system.

A more refined system of spectral classification has recently been widely used, developed by W. W. Morgan, with P. C. Keenan and other collaborators. In this system, employing spectra at 120 A/mm, a two-parameter set of criteria provides by inspection both spectral type and estimated luminosity. The luminosity is indicated by a suffix ranging from Ia, extremely bright supergiant, to III, normal giant, to V, main-sequence or dwarf star. Thus the designation G2 V represents a star like the Sun, G2 III a giant star of nearly the same temperature, with certain luminosity-sensitive features enhanced.

temperature—very hot, blue | solar yellow | cool red

	WC				R		N	carbon sequence
spectral type —	O	B	A	F	G	K	M	normal
	WN						S	peculiar

Fig. 4 Spectral classification from Henry Draper catalog

Peculiar spectra exist, such as the carbon stars (old name R and N, new type C) and S stars which have about the same temperature as K or M stars. Another very important type of peculiarity has been found associated with stars with magnetic fields (see STELLAR MAGNETIC FIELD).

The percentage of stars of different spectral types in the Henry Draper catalog is roughly as follows: B, 3%; A, 27%; F, 10%; G, 16%; K, 37%; M, 7%; other types are rare. This is a selection of stars by apparent brightness and does not represent their true distribution in space, which from the data in Table 1 heavily favors the late-type M dwarfs.

Temperature and luminosity. The major part of the differences in appearance of stellar spectra is caused by the change in the surface temperature

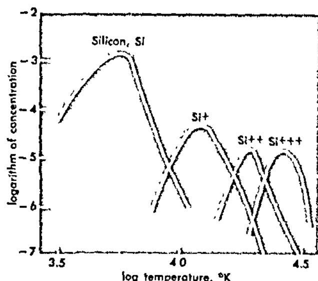


Fig. 5. Effect of temperature on spectral lines of silicon in various states of ionization.

The degree of ionization and excitation of the atoms at a given temperature dictates whether an element will have an appreciable concentration of atoms in the lower atomic level that produces the absorption line. For example, at very high temperature, helium (He) is all He^{++} and thus has no lines, at about 35,000°K, He^{+} dominates, and below 20,000°K, He exists. Below about 12,000°K insufficient atoms of He are excited to the states at 19 electron volts which produce lines in the normal spectral region; no helium lines will be seen (except in the inaccessible ultraviolet) in cooler stars. Stars with He^{+} lines exist and are of type O, He lines occur in type B stars. Stellar temperatures can be accurately determined by quantitative application of these methods, involving the Saha ionization and the Boltzmann excitation equations (Fig. 5).

Absolute magnitude effects occur because the ionization equation depends on pressure, the electron concentration setting the recombination rate. Consequently, stars of low surface gravity, which have lower pressures, will show a given percentage of ionization at a temperature about 500K° lower than those of the main sequence. A red giant of the same temperature as the Sun, 100 times brighter, has 10 times the radius and about 0.01 times the surface gravity (allowing for the larger mass of the giant). The lower gravity and pressure result in an increased level of ionization of sensitive elements. Thus the luminosity classification of a star, a second parameter, is possible after a temperature classification has been made. These effects are calibrated by stars of otherwise known luminosity.

The luminosity of a star is its energy output, either in ergs per second or in units of solar luminosity, or in absolute magnitude. But apparent magnitude and distance in parsecs or parallax must be known. Let a star have luminosity L, radius R, and effective temperature T. Then from Stefan's law and the areas of the star

$$\frac{L_{star}}{L_{\odot}} = \left(\frac{R_{star}}{R_{\odot}}\right)^2 \left(\frac{T_{star}}{T_{\odot}}\right)^4$$

If the temperature corresponding to a given spectral type or color is known, one of a variety of types of Hertzsprung-Russell diagrams can be plotted, connecting luminosity, or absolute magnitude (or apparent magnitudes if all stars of a group are located at the same distance) as ordinates and temperature, or spectral type or color as abscissas. In a diagram in which log L and log T are used, loci of constant radius are straight lines (Fig. 6). Such a diagram, with main sequence and giant branch, serves for the location of many of the various sequences of stars.

Age, evolution, and mass. A group of color-magnitude curves for clusters, galactic (population I) and globular (very old population II), as given by A. R. Sandage and H. C. Arp, shows that, although the fainter end of the main sequence is essentially the same in all groups of stars, the brighter ends vary from one group to another in accordance with differences in age and composition (Fig. 7). Stellar evolution causes such variations.

These H-R diagrams are most significant for the study of the ages, nuclear energy sources, and evolution of the stars. The location of a star in an H-R diagram is completely determined by its mass and chemical composition, if the latter is specified in detail throughout the star.

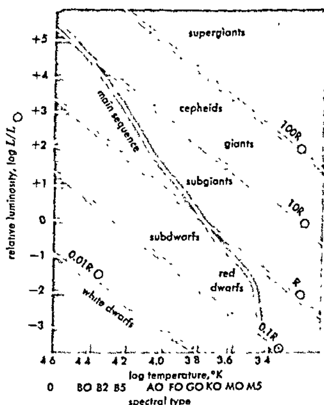


Fig. 6. Main sequence and branch sequences of stars as functions of temperature and relative luminosity. Approximate spectral types are listed below the temperature scale.

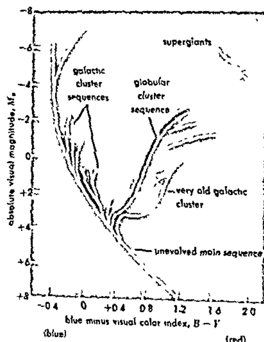


Fig. 7. Color-magnitude diagram for clusters shows variations in brightness for groups of different ages and compositions. Clusters become older from top to bottom of diagram.

A point in an H-R diagram, specified by L and R, does not uniquely determine the mass, however, because of the possible variation of composition with depth. (The latter is caused by the consumption of hydrogen and its conversion to helium by thermonuclear processes.)

The masses of the stars are determined for stars which are members of a double or multiple system, either visual or spectroscopic binaries. The application of Newton's laws provides the mass, although often with considerable uncertainty (see BINARY STARS). If a pair of stars rotates in an orbit whose plane includes the line of sight, so that the star is also an eclipsing binary, and if the lines of both stars are visible in the composite spectrum and show measurable Doppler shifts, it is a two-line spectroscopic binary, so that the size of the orbit, its inclination to the line of sight, the relative masses, and the actual masses are determinable. Such completely observed cases are rare but provide what quantitative information there is about masses, radii and surface temperatures. Visual binaries with well-observed absolute orbits also give masses. The number of favorable circumstances required makes it difficult to measure directly the luminosities, masses, and radii of all the interesting types of stars. Thus many of the desirable relationships between parameters of a star are statistical in nature. The mass-luminosity relation is of this nature.

The stars of the main sequence, for which stellar evolution has not yet been a substantial factor in displacement from their normal positions in an H-R diagram, obey a fairly well-established mass-luminosity relationship (Fig. 8). However, because of stellar evolution, stars move off the main sequence. If they belong to population I, they essentially move horizontally in an H-R diagram, so that they still obey the mass-luminosity relationship. Population II stars with low metal abundances brighten by factors up to a hundred in the subgiant and red giant stage, so that a few cases of clear violation of the mass-luminosity relation are known: for example, the population II visual binary, ζ Herculis A, for which the brighter component is four times brighter than the Sun, although it weighs only 1.07 times as much as the Sun. Other types of stars which deviate from the mass-lum[illegible] the faint [illegible] iables, a [illegible]

Table 3 summarizes surface temperature, luminosity, and radius of typical stars on the main sequence as a function of spectral type and gives, with a good deal less certainty, similar data for giants and supergiants. From the data can be obtained surface gravity $g = GM/R^2$ and mean density $\rho = 3M/4\pi R^3$. A few direct measures of the angular diameters of red supergiants were made by stellar interferometers at the Mount Wilson Observatory and agree approximately with the estimates in Table 3.

Surface temperature. The determination of surface temperature of stars is carried out by various techniques. Measurements of color, because the star's radiation is approximately that of a black body, provide a color temperature. In combination with the theory of stellar atmospheres and the knowledge of the opacity of stellar material, colors yield good measurements of the effective temperatures of stars. Late-type stars with strong molecular bands, however, have highly distorted energy curves and temperatures can be deduced best from infrared and radiometric magnitude, measured with bolometers.

Because all black-body curves have essentially the same shape on the long-wavelength side of

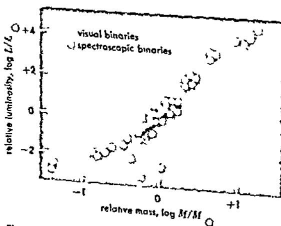


Fig. 8. Mass-luminosity relation. (From observations by K. A. Strand)

Table 3. Approximate physical parameters of the stars

Properties of the Sun

G2, $L = 3.9 \times 10^{33}$ ergs/sec, $M = 2 \times 10^{33}$ g, $T = 5750°K$, $M_v = +4.64$, $R_\odot = 6.96 \times 10^{10}$ cm, $\log g = +4.44$, $\log \rho = +0.16$

Type	Color B − V	M_v	$\log L/L_\odot$	$T/1000°$	$\log R/R_\odot$	$\log M/M_\odot$	Remarks
Main sequence							
O8	−0.3 :*	− 5	+5.05	35	+0.96	+1.25	Uncertain, wide range
B0	−0.32:	− 4.3	+4.66	25	+1.05	+1.15	Uncertain, wide range
B1	−0.28	− 3.5	+4.06	22	+0.85	+1.05	
B2	− .24	− 2.8	+3.72	20	+0.76	+0.95	
B5	− .16	− 1.3	+2.96	15	+ .62	+ .78	
A0	0.00	+ 0.8	+1.73	11	+ .31	+ .45	
A5	+ .19	+ 1.9	+1.21	8.7	+ .24	+ .25	
F0	+ .37	+ 2.5	+0.85	7.6	+ .18	+ .14	
F5	+ .47	+ 3.5	+ .49	6.6	+ .12	+ .08	
G0	+ .60	+ 4.2	+ .21	6.0	+ .06	+ .04	
G5	+ .70	+ 5.2	− .19	5.5	− .06	− .17	
K0	+ .86	+ 6.1	− .55	5.1	− .18	− .22	
K5	+1.24	+ 7.5	− .84	4.4	− .18	− .25	
M0	+1.45:	+ 9.0	−1.23	3.6	− .22	− .30	Uncertain
M2	+1.5 :	+10.0	−1.48	3.2	− .24	− .40	Uncertain
M4	+1.6 :	+12.0	−1.91	3.1	− .42	− .55	Uncertain, wide range
M6	+1.8 :	+15.0	−2.71	2.9	− .76	−0.90	
Giant stars							
G0		+ 0.7	+1.65	5.3	+0.90		Masses uncertain
K0		+ 0.2	+2.04	4.2	+1.30		Masses uncertain
M0		− 0.4	+2.72	3.3	+1.84		Masses uncertain
Supergiant stars							
B0		− 7	+5.66	25	+1.55		All data uncertain
A0		− 7	+4.77	10	+1.90		All data uncertain
K0		− 7	+5.26	3.6	+3.03		All data uncertain
M0		− 7	+5.66	3.0	+3.39		All data uncertain

* Colons indicate discordant determinations

their energy maxima, all hot stars have essentially the same blue color, because their energy maxima are in the vacuum ultraviolet. In addition, hot stars are usually of high luminosity so that interstellar absorption and reddening impedes the analysis of their colors and luminosities.

Strengths of absorption lines of different stages of ionization and excitation may be used to determine the temperature in hot stars. Certain of the hottest O stars, the Wolf-Rayet stars, and the nuclei of planetary nebulae show emission lines excited by a fluorescent conversion of their far ultraviolet into visible radiations. The process is one of photoelectric ionization followed by recombination and emission of subordinate lines. Temperatures for such objects can be determined by the Zanstra method. The range of stellar temperatures is large. The hottest stars have effective temperatures near 50,000–100,000°K; the cooler stars are near 1,500–2,000°K. Although these extremes are somewhat unreliable, the temperatures given in Table 3 serve as a guide to the physical conditions in average stars.

Stellar rotation. Another important property of stars is their rotation on their axes. The prevalence of wide double stars and of close spectroscopic binaries indicates that a large amount of angular momentum is often contained in the material that condensed to form the stars. In close double systems, revolution and rotation are often synchronous. In single stars, especially those of early spectral type, rotation is rapid.

The Sun has an equatorial velocity v of only 2 km/sec. Measurements of the rotational broadening of spectral lines give results contained in Table 4, based largely on the work of A. Slettebak.

Table 4. Mean rotational surface velocities, in km/sec

Type	Dwarfs	Giants
B1–B3	200	127
B5–B7	257	163
B8–A2	177	93
A3–A7	173	202
A9–F2	87	125
F3–F6	31	67
F7–G0	<25	34

The large rotation of the early type-B stars often results in instability, in the form of the ejection of matter from the rapidly rotating equatorial regions. As a result, such stars are often surrounded by disklike rings of low-density material approximately 10 times the radius of the star. These are detected by the presence of emission lines. The rotation drops rapidly down the main sequence and usually cannot be detected in single stars of types later than G0.

Composition. The chemical composition of a star may be deduced from the spectrum of its atmosphere or from the theory of its *internal structure*. There is evidence that stars need not be chemically homogeneous. Products of nuclear reactions may concentrate in the center so that the ratio of helium to hydrogen increases inward. Very slow mixing, however, counteracts the tendency toward complete diffusive separation by gravity of heavy from light elements.

Table 5 **Abundances of elements in normal stars and Sun compared to meteorites and Earth***

Atomic number	Symbol	Element	Logarithm of number of atoms, log N	
			Stellar	Terrestrial
1	H	Hydrogen	12.0	(12.0) †
2	He	Helium	11.1	?
3	Li	Lithium	1.0	3.4
4	Be	Beryllium	2.3	2.7
5	B	Boron	?	2.8
6	C	Carbon	8.3	?
7	N	Nitrogen	8.2	?
8	O	Oxygen	8.9	?
9	F	Fluorine	(6.4)	4.6
10	Ne	Neon	8.6	?
11	Na	Sodium	6.3	6.1
12	Mg	Magnesium	7.5	7.4
13	Al	Aluminum	6.2	6.4
14	Si	Silicon	7.5	7.4
15	P	Phosphorus	(5.3)	5.4
16	S	Sulfur	(7.3)	7.0
17	Cl	Chlorine	(7.2)	5.0
18	Ar	Argon	(6.8)	6.6
19	K	Potassium	5.0	4.9
20	Ca	Calcium	6.1	6.1
21	Sc	Scandium	3.3	2.9
22	Ti	Titanium	4.9	4.8
23	V	Vanadium	4.0	3.8
24	Cr	Chromium	(5.5)	3.3
25	Mn	Manganese	5.0	3.3
26	Fe	Iron	6.9	7.2
27	Co	Cobalt	4.8	4.7
28	Ni	Nickel	5.9	5.9
29	Cu	Copper	4.9	3.8
30	Zn	Zinc	4.5	1.1
31	Ga	Gallium	(2.1)	2.5
32	Ge	Germanium	(3.2)	3.1
37	Rb	Rubidium	(2.3)	2.2
38	Sr	Strontium	2.8	2.7
39	Y	Yttrium	(2.8)	2.4
40	Zr	Zirconium	2.2	3.2
41	Nb	Niobium	1.9	1.4
42	Mo	Molybdenum	1.9	1.8
44	Ru	Ruthenium	1.5	1.6
45	Rh	Rhodium	(−0.1)	0.8
46	Pd	Palladium	(1.0)	+1.3
47	Ag	Silver	(0.2)	0.9
48	Cd	Cadmium	(1.8)	1.4
49	In	Indium	(0.6)	0.5
50	Sn	Tin	(0.6)	1.6
51	Sb	Antimony	(1.7)	0.8
56	Ba	Barium	2.2	2.0
70	Yb	Ytterbium	(1.3)	0.8

* Stellar includes solar and stellar values Terrestrial is Earth or meteorites Parentheses indicate poor values colons added indicate [illegible]
earths have be [illegible]
Pb is very dise [illegible]
values cannot [illegible]
cated by ? bee [illegible]
† H has bee [illegible]

Detailed studies of the composition of stars can only be made in their atmospheres. The ratio of hydrogen to heavy elements is about 8,000–12,000 to 1, by number of atoms, for the Sun and young population I stars. For extreme high-velocity stars, old objects of population II, the abundance of the metals may be 10–100 times lower. Certain elements are unobservable in the spectra of stars and the abundances of heavy and generally rare elements, as well as the isotopic abundances, are best obtained from the crust of Earth or meteorites. Table 5 is a composite resume of current determinations of the abundances. Comparison with Earth, for the metals, shows good agreement. In the process of formation of Earth most of its H, He, C, N, O, Ne, and other normally light gaseous elements were lost, because of low atomic weight, the only atoms retained being those in chemical combination Therefore, good agreement cannot be expected Certain features of the terrestrial abundance curve, notably the abundance ratio of silicon to iron are not accurately known and are estimated from meteorites so that some gross adjustments may still be needed (*see* ELEMENTS, GEOCHEMICAL DISTRIBUTION)

Stellar or solar abundances are determined by a set of measurements and interpreted by theory. Wavelengths of lines yield identification of the elements, and line intensities can be interpreted in terms of the numbers of atoms in the atomic levels producing the lines. However, to do this, the transition probability in the line must be known, either from quantum mechanics or laboratory measurements In addition, corrections for the state of ionization and excitation must be applied to permit computation of the total concentration of the element, and these corrections depend strongly on the temperature This is especially true for atoms whose visible lines arise from levels of high excitation potential, for example, in the Sun only one atom of helium in 10^{20} is likely to be excited into a state capable of producing a visible absorption line [illegible] nations are [illegible] which pre- [illegible] far ultraviolet spectra, where the resonance lines of many important elements are located The subject of stellar atmospheres is an important branch of modern astrophysics. It relies heavily on the mathematical solution of problems of the diffusion of light outward through an absorbing and emitting medium and on physical information about the shape of atomic absorption lines, the amount of continuous absorption of light (which limits the depth to which we can see), and the theories of line broadening

Motions of stars. Galactic dynamics and kinematics are an important part of the subject of stellar statistics and are intimately connected with the distribution of the stars in space (*see* GALAXY, EXTERNAL; GALAXY, THE) Our galaxy has a mass of about 2×10^{11} Suns of which only a small fraction is visible from Earth These stars move in orbits

Table 3. Approximate physical parameters of the stars

Properties of the Sun

G2, $L = 3.9 \times 10^{33}$ ergs/sec, $M = 2 \times 10^{33}$ g, $T = 5750°K$, $M_v = +4.64$, $R_\odot = 6.96 \times 10^{10}$ cm, $\log g = +4.44$, $\log \rho = +0.16$

Type	Color B − V	M_v	$\log L/L_\odot$	$T/1000°$	$\log R/R_\odot$	$\log M/M_\odot$	Remarks
Main sequence							
O8	−0.3 :*	−5	+5.05	35	+0.96	+1.25	Uncertain, wide range
B0	−0.32:	−4.3	+4.66	25	+1.05	+1.15	Uncertain, wide range
B1	−0.28	−3.5	+4.06	22	+0.85	+1.05	
B2	−.24	−2.8	+3.72	20	+0.76	+0.95	
B5	−.16	−1.3	+2.96	15	+.62	+.78	
A0	0.00	+0.8	+1.73	11	+.31	+.45	
A5	+.19	+1.9	+1.21	8.7	+.24	+.25	
F0	+.37	+2.5	+0.85	7.6	+.18	+.14	
F5	+.47	+3.5	+.49	6.6	+.12	+.08	
G0	+.60	+4.2	+.21	6.0	+.06	+.04	
G5	+.70	+5.2	−.19	5.5	−.06	−.17	
K0	+.86	+6.1	−.55	5.1	−.18	−.22	
K5	+1.24	+7.5	−.84	4.4	−.18	−.25	
M0	+1.45:	+9.0	−1.23	3.6	−.22	−.30	Uncertain
M2	+1.5 :	+10.0	−1.48	3.2	−.24	−.40	Uncertain
M4	+1.6 :	+12.0	−1.91	3.1	−.42	−.55	Uncertain, wide range
M6	+1.8 :	+15.0	−2.71	2.9	−.76	−0.90	
Giant stars							
G0		+0.7	+1.65	5.3	+0.90		Masses uncertain
K0		+0.2	+2.04	4.2	+1.30		Masses uncertain
M0		−0.4	+2.72	3.3	+1.84		Masses uncertain
Supergiant stars							
B0		−7	+5.66	25	+1.55		All data uncertain
A0		−7	+4.77	10	+1.90		All data uncertain
K0		−7	+5.26	3.6	+3.03		All data uncertain
M0		−7	+5.66	3.0	+3.39		All data uncertain

* Colons indicate discordant determinations

their energy maxima, all hot stars have essentially the same blue color, because their energy maxima are in the vacuum ultraviolet. In addition, hot stars are usually of high luminosity so that interstellar absorption and reddening impedes the analysis of their colors and luminosities.

Strengths of absorption lines of different stages of ionization and excitation may be used to determine the temperature in hot stars. Certain of the hottest O stars, the Wolf-Rayet stars, and the nuclei of planetary nebulae show emission lines excited by a fluorescent conversion of their far ultraviolet into visible radiations. The process is one of photoelectric ionization followed by recombination and emission of subordinate lines. Temperatures for such objects can be determined by the Zanstra method. The range of stellar temperatures is large. The hottest stars have effective temperatures near 50,000–100,000°K; the cooler stars are near 1,500–2,000°K. Although these extremes are somewhat unreliable, the temperatures given in Table 3 serve as a guide to the physical conditions in average stars.

Stellar rotation. Another important property of stars is their rotation on their axes. The prevalence of wide double stars and of close spectroscopic binaries indicates that a large amount of angular momentum is often contained in the material that condensed to form the stars. In close double systems, revolution and rotation are often synchronous. In single stars, especially those of early spectral type, rotation is rapid.

The Sun has an equatorial velocity v of only 2 km/sec. Measurements of the rotational broadening of spectral lines give results contained in Table 4, based largely on the work of A. Slettebak.

Table 4. Mean rotational surface velocities, in km/sec

Type	Dwarfs	Giants
B1–B3	200	127
B5–B7	257	163
B8–A2	177	93
A3–A7	173	202
A9–F2	87	125
F3–F6	31	67
F7–G0	<25	34

The large rotation of the early type-B stars often results in instability, in the form of the ejection of matter from the rapidly rotating equatorial regions. As a result, such stars are often surrounded by disklike rings of low-density material approximately 10 times the radius of the star. These are detected by the presence of emission lines. The rotation drops rapidly down the main sequence and usually cannot be detected in single stars of types later than G0.

The high-velocity population II stars are moving in galactocentric orbits of high eccentricity and inclination to the galactic plane. If they were moving with a velocity vector measured with respect to the local standard of rest in the same direction as the galactic rotation, their kinetic energy would be so great compared to the gravitational force of the galaxy as to exceed escape velocity. Thus it is found that stars with large velocity vectors, greater than 65 km/sec with respect to the Sun, are absent in about one quadrant of the sky, in the forward direction of galactic rotation.

Solar motion. The solar motion is defined by a vector giving the direction of A, the apex of solar motion, and U, the velocity with respect to the local standard of rest. The radial velocity of the Sun with respect to a direction located at angular distance λ from the apex is a differential motion, $\delta V = -U \sin \lambda$ By averaging over the observed radial velocities of stars of a given type, at moderate distances from the Sun (or correcting separately for galactic rotation if necessary), both U and λ can be determined

The apex of solar motion with respect to the nearby population I stars is at right ascension 18^h, declination $+29°$, with U about 20 km/sec With respect to the population II stars, the direction of the apex is strongly dependent on the velocity dispersion

Proper motion The transverse motions of the stars with respect to a fixed coordinate system on the plane of the sky are seen as proper motions, which are usually given in seconds of arc per year. The total proper motion μ is derived from the component μ_α, in right ascension, in seconds of time, and μ_δ, in declination, in seconds of arc, by

$$\mu^2 = (15\,\mu_\alpha \cos \delta)^2 + \mu_\delta^2$$

Proper motion of a star is measured by displacements on photographic plates taken at a sufficiently long interval of time The standard of rest is usually established by stars of accurately known absolute positions and motions. In principle, an ideal standard of rest would be provided by the faint extragalactic nebulae and work is in progress with the latter technique. However, normally, absolute positions of reference stars are determined by visual observations, accurately timed, of the transits of stars, using *meridian* circles About 200 stars have motions ...

... of a star as observed from the Sun form a vector diagram (Fig 10). If, in a year the star moves from S to S' at velocity v, the velocity in the line of sight is radial velocity V, and the velocity across the line of sight is proper motion μ, in seconds of arc per year Parallax p is also expressed in seconds of arc; then tangential velocity T is $T = 4.74\ \mu/p$, the space velocity v is $v^2 = V^2 + T^2$, and the radial and tangential velocities are $V = v \cos \theta$ and $T = v \sin \theta$ All velocities are expressed in km/sec.

Conversely if the stars were all standing still and the Sun alone were moving, or if the stars' motions are averaged out, the formula above for the tangential velocity indicates that a star would show a proper motion due to the projection of the solar motion on the plane of the sky. This permits the determination of the solar motion from the proper motions of nearby stars for which p is not too small. Similarly, if the observed motions of a group of stars, after correction for solar motion, are used, an estimate is obtained of the average parallax. Special catalogs give accurate positions and proper motions.

Radial velocity Spectroscopic measures of Doppler shifts give radial velocity. A general catalog of stellar radial velocities of 15,000 stars is available. Stellar radial velocities are determined with accuracies up to ±0.1 km/sec; the largest velocities are about 400 km/sec, but only 4% are greater than 60 km/sec.

Galactic dynamics has as its main result the interpretation of the space motions of stars in terms of galactic rotation, orbits, and the distribution of mass in our galaxy. From the rotational-velocity curve for circular orbits, the mass is determined; this method has been applied to a few extragalactic nebulae with the general result that masses of galaxies range from 10^9 to 2×10^{11} Suns The dynamics of such features as the spiral arms have not yet been understood

Spatial distribution. The distribution of stars in space is the subject of studies of stellar statistics and galactic structure The Milky Way is the dominant feature of our galaxy, even to the eye, and represents the mean plane in which the stars are concentrated It is nearly a great circle, indicating that the Sun lies near the galactic plane.

The number of stars at a given apparent magnitude is a function of galactic latitude and longitude; in the galactic plane the complex structure of interstellar clouds of dust that absorb light and the irregularities of spiral structure produce a somewhat irregular, patchy appearance, with a maximum number of faint stars in the direction of the galactic center in *Sagittarius*, with subsidiary maxima in Cygnus and Carina. The latter are probably caused ...

Fo ... latitude ... which gives $\log N_{m,b}$ where $N_{m,b}$ is the number of stars brighter than apparent magnitude m, per square degree, at latitude b. The light of the entire sky is equivalent to that of one star of magnitude −6.6 The number of stars increases very rapidly with m, about three-

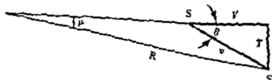


Fig 10 Components of proper and radial motions of

around their common center of gravitation, located about 8500 parsecs from Earth in the constellation Sagittarius. The galaxy is highly flattened by its systematic rotation, the linear rotational velocity being approximately 220 km/sec at the Sun's distance from the galactic center; this corresponds to a rotation period of 200,000,000 years.

Differential rotation. Within the galaxy, the rotation is described with respect to an external stationary frame of reference, from which it would also appear as a differential rotation, with angular velocity varying with distance from the galactic center (Fig. 9). This rotation curve is established by study of the relative velocities of distant stars, and of clouds of interstellar hydrogen seen by their 21-cm radio-frequency radiation, both measured with respect to the Sun. Were the galaxy to rotate as a rigid body, there would be no relative velocities of approach or recession; that is, no radial velocity of a systematic character. A differential rotation, however, is detectable in the transverse motions of the stars, with respect to an outside frame of reference, even in the rigid body case. In the nonrigid body rotation, the distant stars show systematic velocities in certain preferred directions. The relative radial velocity $V(r,l)$ is

$$V(r,l) = R_0[\omega(R) - \omega(R_0)] \sin (l - l_0)$$

where l is the azimuthal coordinate, in the galactic plane, of the star, l_0 is that of the galactic center, as seen from the Sun, R_0 is the Sun's distance to the galactic center, and R that of the star; the motion is assumed to be in the galactic plane. To a first approximation, for small distances r measured from Sun to star, the above formula can be written as

$$V(r,l) = 2rA \sin 2(l - l_0)$$

where A is the first-order galactic rotation constant, which lies between about 13 and 17 km/sec per 1000 parsecs. This double wave is visible in the mean radial velocities of the stars in the range of distance 300 to 3000 parsecs; at larger distances higher-order terms must be included

Peculiar velocity. Superposed on the systematic rotation of the galaxy are individual motions of the stars. Each star moves in a somewhat elliptical orbit and therefore shows a peculiar velocity with respect to the local standard of rest, the standard moving in a circular orbit around the galactic center. The Sun has an orbit of small ellipticity and inclination, so that solar motion with respect to the mean of neighboring stars can be detected by analysis of either radial velocities or proper motions of nearby stars.

Stars can be considered to be particles in a gas which has no collisions, with gravitational distant encounters between pairs of stars only slightly altering their orbits The velocity dispersion of the stars with respect to their local standard of rest is essentially Maxwellian, except that instead of spherical symmetry, the stars display greater mobility, called preferential motion, in the directions toward and away from the galactic center than they do in other directions; that is, their velocity distribution is ellipsoidally symmetric.

Escape velocity Superposed on this is an asymmetry of motion found for stars of high peculiar velocity The velocity distribution found for the stars is not a single Maxwellian one, but has essentially at least two different dispersions. One group is characterized by a small dispersion of space motion with respect to the local standard of rest, for example about ±8 km/sec for B stars, increasing to ±20 km/sec for M stars of population I

Another group, the population II stars, have velocity dispersions ranging from ±30 to ±150 km/sec

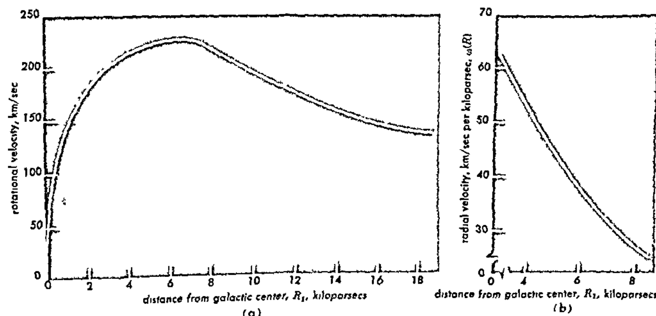


Fig. 9. (a) Rotational velocity varies with distance ... velocity would be constant instead of decreasing with radius as it does.

Scutum, as illustrated The stars in such clouds may appear unevenly distributed because of the presence of obscuring interstellar dust and gas. [illegible] galaxies to our own. [H.S.H.]

Star clusters

Groups of stars held together by gravitational attraction The two chief types are galactic or open clusters, containing from a dozen up to many hundred stars, and globular clusters, composed of thousands to hundreds of thousands of stars A relative of the star cluster is the stellar association, a group of dozens or hundreds of stars spread loosely over a larger volume of space. Star clusters are important in outlining the shape and extent of our galaxy and in deriving theories of stellar evolution on the assumption that stars of a given cluster were formed at the same time

Galactic clusters. Open clusters lie along the backbone of our galaxy, being strongly concentrated to the central plane of the Milky Way A dozen are visible to the unaided eye, over 500 are cataloged, and many more must exist. Most galactic clusters have an asymmetrical appearance (Fig. 1).

Distances and dimensions. The distances to galactic clusters range from 40 parsecs for the [illegible]

geometric methods, measures are made of trigonometric parallaxes or of stellar motions. By photometric methods the apparent and absolute magnitudes of the stars are determined See PARALLAX (ASTRONOMY).

Angular diameters of galactic clusters range from several degrees down to several minutes of arc, the corresponding linear diameters ranging from 15 parsecs down to 2 parsecs. From a study of the way in which linear diameters appeared to increase with increasing distance, the absorption of light in space was deduced by R J Trumpler in 1930.

Spectral characteristics The brightest stars in some clusters, like the Pleiades, are blue, of spectral type B; in others like the Hyades or Praesepe, they are yellow or red. Stars with luminosities brighter than absolute magnitude −3 are found, and sometimes supergiants up to −7.

Most stars fall along the highly populated branch of the spectrum-luminosity diagram known as the main sequence. The point where this sequence starts furnishes a criterion of the age of the cluster. Galactic clusters show great diversity in their spectrum-luminosity diagrams. They may be classified on this basis, as well as by richness and central concentration They may contain such types of stars as bright O stars visual and spectroscopic binaries certain kinds of variables, and white dwarfs. Some clusters, like the Pleiades, contain amounts of nebulosity equivalent to many solar masses

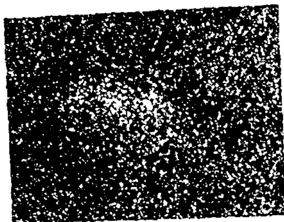


Fig. 1 The double cluster in Perseus. (Yerkes Observatory photograph)

Motions. Measures of proper motion and radial velocity show the cluster stars to be sharing a common motion in space, with velocities up to tens of kilometers per second Galactic clusters whose most prominent characteristic is a large common proper motion of the stars are called moving clusters. The diagram of proper motions in a cluster shows a conspicuous convergent point, where their parallel motions appear to meet in space This category includes the Taurus and Ursa Major moving clusters

Age and dissolution By comparing the spectrum-luminosity diagram with a standard main-sequence, the age of a cluster may be determined At one extreme are young clusters, formed in recent geologic times, like NGC 2244 and 2264 with ages of the order of 10^6 years. At the other extreme is Messier 67, with an age of 5×10^9 years, comparable with old systems like globular clusters The lifetime of a cluster will depend on its mass The more massive will lose fewer stars by "evaporation" from the cluster and by disruption from interstellar clouds and galactic tidal forces

Stellar associations. Systems where early type (O to B2) stars are more numerous than in the surrounding field are cataloged as stellar associations About 50 are cataloged The radii range up to 200 parsecs They are perishable, lasting 10,000,000–20,000,000 years

Globular clusters. Groups of thousands to hundreds of thousands of stars in globular symmetry constitute globular clusters Though they are scattered widely in galactic latitude, their strong concentration toward the region of Sagittarius-Scorpio led Harlow Shapley in 1917 to postulate this as the center of our galaxy. Several are visible to the unaided eye, like Messier 13, the great cluster in Hercules (Fig 2) A total of 118 have been cataloged in our galaxy, including several so far distant that they are really intergalactic [illegible]

Table 6 Apparent distribution log $N_{m,b}$ of stars as a function of galactic latitude

m \ b	0°	10°	25°	50°	90°
6	−0.89	−0.97	−1.16	−1.35	−1.43
8	0.00	−0.08	−0.26	−0.45	−0.56
10	+0.89	+0.79	+0.59	+0.40	+0.26
12	+1.74	+1.63	+1.41	+1.18	+1.00
14	+2.57	+2.43	+2.17	+1.88	+1.65
16	+3.33	+3.19	+2.84	+2.48	+2.21
18	+4.01	+3.87	+3.42	+2.98	+2.68
20	+4.60	+4.46	+3.90	+3.38	+3.07

fold per magnitude, in the galactic plane; the rate of increase is slower for fainter stars and at higher galactic latitudes. The galactic concentration of stars increases for fainter stars, a natural result because the faint stars are on the average more distant.

Certain types of objects are highly concentrated toward the galactic plane, particularly cepheid variables, luminous O and B stars, and galactic clusters. Others like M dwarfs and long-period variables are less concentrated, while globular clusters and RR Lyrae variable stars show almost no concentration.

Part of this effect is caused by luminosity differences; for example, M dwarfs are so intrinsically faint that they cannot be seen to great distances On the other hand O and B stars are intrinsically highly concentrated to the galactic plane while the globular clusters and other extreme population II high-velocity stars are found at great heights, up to 15,000 parsecs from the plane

In the galactic plane the frequency distribution of the types of stars varies from point to point. Interstellar gas and dust is highly concentrated in the spiral arms, together with O and B stars and other population I objects of high luminosity In the neighborhood of the Sun, the luminosity function given in Table 7 is a useful average value It gives the density of stars per cubic parsec as a function of absolute visual magnitude $\phi(M)$ within range $M + ½$ to $M - ½$ However, outside the spiral arms, and at heights greater than 100 parsecs from the galactic plane, the high-luminosity end of $\phi(M)$ is cut off

At great heights, stars with M_v less than +3 are rare The space density of all types of stars together is about 0.1 solar masses per cubic parsec in the galactic plane and decreases rapidly with height above the plane, following an approximately exponential law, with a scale height h_0, which varies with the type of star. The B stars and the interstellar gas clouds have h_0 about 100 parsecs; RR Lyrae variables have h_0 in the order of 2500 parsecs. This difference reflects the different kinetic energies of slow and fast moving stars The space density of stars increases greatly towards the center of the galaxy, possibly by a factor of 100 The thickness of the galaxy also increases; the galactic center is composed mainly of stars of population II, which show a very large h_0. [J.L.GR.]

Table 7. Frequency log $\phi(M)$ of stars in the neighborhood of the Sun, as a function of luminosity

M	log $\phi(M)$	M	log $\phi(M)$	M	log $\phi(M)$
−6	−7.90	0	−4.32	+ 6	−2.51
−5	−6.93	+1	−3.66	+ 8	−2.51
−4	−6.35	+2	−3.23	+10	−2.36
−3	−5.75	+3	−3.14	+12	−2.03
−2	−5.25	+4	−2.81	+14	−1.94
−1	−4.93	+5	−2.65		

Bibliography: L. H. Aller, *Astrophysics: The Atmospheres of the Sun and Stars*, 1953; A. Becvar, *Skalnate Pleso Atlas of the Heavens*, 1950; B Boss, *General Catalogue of 33342 Stars*, 1937, E. M. Burbidge, G. R. Burbidge, W. A. Fowler, and F. Hoyle, *Rev. Mod. Phys*, 29:547–650, 1957; A. J. Cannon and E C Pickering, *The Henry Draper Catalogue of Stellar Spectra*, Harvard Obs Ann, vols 91–100, 1918–1936; S. Fluegge (ed.), *Handbuch der Physik*, vol. 50, 1957, vol 51, 1958; J. L. Greenstein (ed), *Stellar Spectra*, 1960; L. F. Jenkins, *General Catalogue of Trigonometric Stellar Parallaxes*, 1952, H H Landoldt, *Zahlenwerte und Funktionen aus Physik Chemie, Astronomie, Geophysik und Technik*, vol. 3, 1952, W. W. Morgan, P. C Keenan, and E. Kellman, *An Atlas of Stellar Spectra*, 1943, F. Schlesinger and L. F. Jenkins, *Catalogue of Bright Stars*, 2d ed., 1940; H. E. Suess and H. C. Urey, *Rev. Mod. Phys*, 28:53–74, 1956, R. E Wilson, *General Catalogue of Stellar Radial Velocities*, Carnegie Inst. Washington Publ 601, 1953.

Star clouds

Aggregations of thousands or millions of stars spread over hundreds or thousands of light years in space The Milky Way is composed of such star clouds, the heaviest clouds being in the richest parts such as Cygnus, Sagittarius, Carina, and



The great star cloud in Scutum (Yerkes Observatory photograph)

Scutum, as illustrated. The stars in such clouds may appear unevenly distributed because of the presence of obscuring interstellar dust and gas.

The term has also been applied to the Large and Small Magellanic Clouds which appear similar to Milky Way clouds but are actually the nearest galaxies to our own. [H.S.H.]

Star clusters

Groups of stars held together by gravitational attraction. The two chief types are galactic or open clusters, containing from a dozen up to many hundred stars, and globular clusters, composed of thousands to hundreds of thousands of stars. A relative of the star cluster is the *stellar association*, a group of dozens or hundreds of stars spread loosely over a larger volume of space. Star clusters are important in outlining the shape and extent of our galaxy and in deriving theories of stellar evolution on the assumption that stars of a given cluster were formed at the same time.

Galactic clusters. Open clusters lie along the backbone of our galaxy, being strongly concentrated to the central plane of the Milky Way. A dozen are visible to the unaided eye, over 500 are cataloged and many more must exist. Most galactic clusters have an asymmetrical appearance (Fig. 1).

Distances and dimensions. The distances to galactic clusters range from 40 parsecs for the Hyades up to 5000 parsecs for faint clusters. Those more distant are difficult to detect against a rich star background. For distance determination by geometric methods, measures are made of trigonometric parallaxes or of stellar motions. By photometric methods the apparent and absolute magnitudes of the stars are determined. *See* PARALLAX (ASTRONOMY).

Angular diameters of galactic clusters range from several degrees down to several minutes of arc, the corresponding linear diameters ranging from 15 parsecs down to 2 parsecs. From a study of the way in which linear diameters appeared to increase with increasing distance, the absorption of light in space was deduced by R. J. Trumpler in 1930.

Spectral characteristics. The brightest stars in some clusters, like the Pleiades, are blue, of spectral type B, in others like the Hyades or Praesepe, they are yellow or red. Stars with luminosities brighter than absolute magnitude -3 are found, and sometimes supergiants up to -7.

Most stars fall along the highly populated branch of the spectrum-luminosity diagram known as the main sequence. The point where this sequence starts furnishes a criterion of the age of the cluster. Galactic clusters show great diversity in their spectrum-luminosity diagrams. They may be classified on this basis, as well as by richness and central concentration. They may contain such types of stars as bright O stars, visual and spectroscopic binaries, certain kinds of variables, and white dwarfs. Some clusters, like the Pleiades, con-

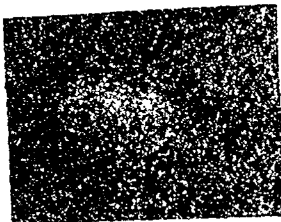


Fig. 1. The double cluster in Perseus. (Yerkes Observatory photograph)

tain amounts of nebulosity equivalent to many solar masses.

Motions. Measures of proper motion and radial velocity show the cluster stars to be sharing a common motion in space, with velocities up to tens of kilometers per second. Galactic clusters whose most prominent characteristic is a large common proper motion of the stars are called moving clusters. The diagram of proper motions in a cluster shows a conspicuous convergent point, where their parallel motions appear to meet in space. This category includes the Taurus and Ursa Major moving clusters.

Age and dissolution. By comparing the spectrum-luminosity diagram with a standard main-sequence, the age of a cluster may be determined. At one extreme are young clusters, formed in recent geologic times, like NGC 2244 and 2264 with ages of the order of 10^6 years. At the other extreme is Messier 67, with an age of 5×10^9 years, comparable with old systems like globular clusters. The lifetime of a cluster will depend on its mass. The more massive will lose fewer stars by "evaporation" from the cluster and by disruption from interstellar clouds and galactic tidal forces.

Stellar associations. Systems where early type (O to B2) stars are more numerous than in the surrounding field are cataloged as stellar associations. About 50 are cataloged. The radii range up to 200 parsecs. They are perishable, lasting 10,000,000–20,000,000 years.

Globular clusters. Groups of thousands to hundreds of thousands of stars in globular symmetr-

... Shapley in 1917 to postulate this as the center of our galaxy. Several are visible to the unaided eye, like Messier 13, the great cluster in Hercules (Fig. 2). A total of 118 have been cataloged in our galaxy, including several so far distant that they are really intergalactic ...

est to 40,000 for distant, and 130,000 for intergalactic. They are too great for geometric methods of distance determination, but photometric methods involving color-magnitude diagrams and RR Lyrae stars can be used The apparent diameters range from 65′ to 1′, and the linear diameters from 190 to 20 parsecs.

Structure. Globular clusters differ markedly in their degree of central concentration and are classified on this basis. A few are noticeably elliptical In some, the frequency of stars falls off as the cube of the distance from the center A count of 44,500 stars in the typical cluster Messier 3, with the Mt. Palomar 200-in. telescope, shows that 95% of the visual light of the cluster comes from stars intrinsically brighter than our Sun, while 90% of the mass is contributed by fainter stars. The estimated mass is 2.45×10^5 solar masses.

Density of stars near the center is high, 50 stars per cubic parsec compared with one star per 10 cubic parsecs near the Sun An average cluster density is one star per 2 cubic parsecs. The massive clusters Omega Centauri and 47 Tucanae have higher densities; at the center of the latter the starlight would be equivalent to several thousand full Moons.

Color-magnitude diagrams differ appreciably from those of galactic clusters. Stars of absolute magnitude brighter than −3 are absent The brightest stars are yellow-red; the main sequence is represented by large numbers of stars from type F down, with a horizontal branch near absolute magnitude 0. The spectra show peculiarities attributed to old stars.

Variable stars abound in some clusters. Messier 3, the richest, has 187, and most globulars contain some. Short-period, RR Lyrae stars make up 90% of the variables

Motions The symmetry of a globular cluster is doubtless caused by rotation about its central mass but the great distances of the clusters render determination of individual star motions difficult Radial velocities measured for 50 clusters range from −360 km/sec to +290 km/sec. These large values are a combination of the galactic rotation of the Sun and of the cluster. Clusters move around the galactic center in orbits with semimajor axes up to 25,000 parsecs.

Age, formation, and dissolution. Evolutionary models give an estimate of the time taken for stars to evolve to certain spectrum-luminosity characteristics. The ages of stars so inferred in globular clusters (and therefore of the clusters) are about $6 \pm 2 \times 10^9$ years. Globular clusters may have formed from dense knots of diffuse material in various shells at specific distances from the galactic center over a period of 1,000,000,000 years Dynamically these clusters are so stable that their individual stars will die as stars before the clusters disintegrate.

Clusters in extragalactic systems. Star clusters comparable to those belonging to our galaxy have been observed in more than a dozen external galaxies. It is difficult to distinguish the type of cluster in such distant objects. Both types have been found in the Large and Small Magellanic Clouds and Messier 31 in Andromeda. Particularly rich in globular clusters is Messier 87, a giant elliptical galaxy in Virgo, with more than 1000 clusters. *See* GALAXY, THE; HYADES; PLEIADES. [H.S H.]

Bibliography: H S Hogg, Star clusters, in S. Fluegge (ed), *Handbuch der Physik*, vol 53, 1959.

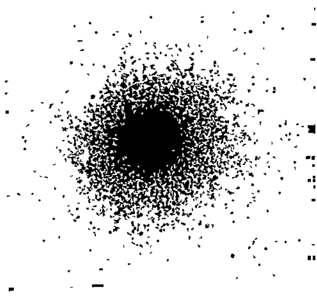


Fig 2. The great globular cluster in Hercules, Messier 13 Photographed with 200-in. telescope. (*Mount Wilson and Palomar Observatories*)

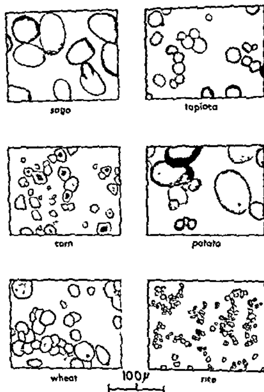
Starch

The reserve carbohydrate stored usually in the seeds, roots, or stems of plants It is second only to cellulose in abundance as a source of carbohydrates. For many centuries, corn, wheat, potato, and rice have supplied the basic carbohydrate needs of mankind. *See* CARBOHYDRATE

Although starch is widespread in plants, only a few sources are sufficiently abundant to make the extraction of the starch commercially feasible These sources are corn, tapioca, potato, sago, waxy maize, wheat, sorghum, rice, and arrowroot The naturally occurring starch is separated from the seed, as in corn, wheat, waxy maize, sorghum, and rice; from the root, as in tapioca, potato, and arrowroot; or from the stem, as in sago, by a variety of methods The usual procedure consists in cleaning the plant material, which is then ground, soaked, washed sieved, and filtered The washed starch is recovered as a filter cake before it is dried and ground

Starch in this form is a white powder, and the unaided eye can detect little difference between the various starches extracted from different plants. All the starches are insoluble in alcohol, most solvents, and cold water A dilute solution of iodine stains starch a blue to bluish-red color

Under the microscope, the starches appear as cellular or granular material with shapes varying



Microscopic appearances of starch (National Starch Products, Inc.)

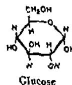
with the botanical origin. A few of these microscopic appearances are shown in the illustration. The granules vary in size from about 2 to 100 μ. Thus each plant builds a starch granule to suit its own purpose.

Molecular structure. Granular differences are related to the basic differences in molecular structures which the various plants build for their individual needs. Starch and cellulose, the two principal carbohydrates of plant origin, are very similar in molecular structure. Both are natural polymers of glucose. The glucose units are joined by β-1,4-glucoside linkages to form cellulose straight chained molecules. α-1,4-glucoside linkages bond the straight-chained starch polymers together. In addition to straight-chained starch molecules known as amylose, there are branched starch molecules, known as amylopectin. The branches are formed occasionally in treelike fashion from the 1,6-glucoside linkage in a normal 1,4-glucoside straight-chained polymer. *See* GLUCOSE.

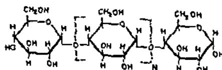
Thus there are two basic types of starch molecules: the linear starch polymer and the branched starch polymer. Amylose molecules vary in size from about 100 to over 1000 glucose units. Amylopectin molecules are larger by three or more times. The sizes of the molecules and the amounts of each type present in the starch granule determine the specific properties of the individual starch. Both types are normally synthesized by and are present in plants. Most starch granules, as produced by nature, contain approximately 20 to 30% of the amylose type molecule, and the balance of the amylopectin type. Starches from the so-called waxy grains contain only amylopectin molecules. Some special hybrids produce granules high in amylose content.

Forms of processed starch. The starch granules are formed by attractive forces between these large carbohydrate molecules. The linear portions of the molecules tend to associate together into micelles ... molecules together into a ... cture is fairly ... lowever, when ... tarch in water is increased to a critical point, called the gelatinization temperature, water penetrates the granules to hydrate and swell them to produce a viscous mass. Gelatinization temperatures vary (60–75°C) from starch to starch. Further swelling, with more thickening, occurs as the temperature is increased, and the starch is cooked. Starches lose their unique microscopic appearance or shape as gelatinization proceeds and they become fully cooked. The characteristics of the viscous solutions vary from starch to starch. The root-type starch solutions, after cooling to room temperature, are clearer, more fluid, and cohesive in texture, while the cereal type starches produce cloudy, less fluid starch pastes that tend to be jellylike in texture.

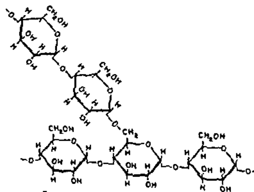
The natural starches find a variety of uses as thickeners in foods and industrial processes. Many times, however, these starches are processed further to obtain viscosity or texture differences of even



Glucose



Linear amylose starch molecule



Branched amylopectin starch molecule

greater use. For example, starch may be converted with enzymes (amylases) to produce lower-viscosity types for sizing purposes. In the presence of acid, and at temperatures below the gelatinization point, starch may be converted in water suspension by breaking some of the molecules to obtain starches which give less viscous solutions. These thin boiling-type starches find many uses where increased concentration is desirable without more viscosity. These cooked starch pastes usually set on cooling with increased jellylike textures The molecular size of starch may also be reduced with oxidizing agents. Carboxyl groups are introduced into the molecules in this way, and cooked starch pastes of these types are less jellylike in texture.

When starch is roasted in dry form, usually in the presence of a small amount of acid, a wide variety of starch products is obtained. The color usually darkens to shades of tan to brown, and the viscosity is reduced with increased roasting. This dextrinization process is a complex chemical reaction, and consists of a combination of hydrolysis, oxidation, and transglucosidation of the starch molecule, as well as repolymerization of some of the reaction products The normally insoluble (in cold water) starch granule is converted with increasing dextrinization into smaller molecules which are soluble in cold water. Some common names for dextrins are British gums, white dextrins, and canary or yellow dextrins These products find many industrial uses Concentrated solutions of dextrins are sticky and have wide adhesive applications.

If the starch molecule is depolymerized further, usually by acid hydrolysis, a variety of sugars and dextrose or glucose is obtained. These in syrup or dry form find wide application in food products as sweeteners. *See* CORN.

Carboxyl (COOH) groups are introduced into the starch molecule on oxidation Other groupings or radicals, such as ethers or esters, can be introduced to produce different properties to the individual starch. Starch molecules also may be joined to form larger molecules Thus, a wide variety of starch properties may be obtained to suit the needs of industry. *See* FOOD ENGINEERING. [T A W]

Bibliography: F H. Frost et al., *Starch and Starch Products in Paper Coating*, Tech. Assoc. Pulp Paper Ind., Monograph Ser., 17, 1957; R. W. Kerr, *Chemistry and Industry of Starch*, 2d ed., 1950; National Starch Products Inc., *The Story of Starches*, 1953.

Starfish

A member of the class Asteroidea, phylum Echinodermata There are about 200 living species, all marine.

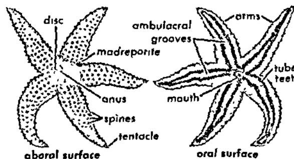
Economic importance. Starfishes are of some economic importance in areas where oysters are produced because they prey heavily upon oysters and can cause serious losses Other than this they are of little importance. Some are used in zoology ... specimens. The early embryology is also studied as an example of holoblastic cleavage, and young stages are frequently utilized in various embryological experiments (*see* EMBRYOLOGY, EXPERIMENTAL; INVERTEBRATE EMBRYOLOGY). Dried specimens are frequently used as decorations.

Structure. Starfishes are radially symmetrical, with movable calcareous skeletons Typically they are 5-armed, but some have a large, 5-sided disk without arms, whereas others may have 4–50 arms. Most starfishes are only a few inches across, but one has a spread up to 32 in.

Starfishes have a water vascular system, unique to the Echinodermata, which is used in locomotion food handling, and respiration. In this system, water is used to fill a ring canal around the disk and branches which run down each arm. The branches connect to the ampullae of the numerous tube feet, which are small suction-cup structures This hydraulic system enables the starfish to exert maximum force in such activities as opening oysters and pulling itself along with minimum muscular strength. The system opens on the dorsal surface of the disk through a pore, the madreporite.

In addition to the tube feet, respiration is accomplished by dermal papulae which are small saclike structures projecting through openings in the skeleton and leading to the coelom. The digestive system extends into the arms The mouth is ventral and connects directly to the centrally located stomach. Starfishes are covered with cilia on both surfaces, which probably aid in keeping their surfaces free of debris, plants, and animals. Most starfishes also have small pincerlike structures, the pedicillariae, on their surfaces, which aid in keeping them clean. The circulatory system is reduced to a ring around the mouth and a small vessel running down each arm A similarly situated nerve ring with four nerves, which extend along each arm, comprises the nervous system

Reproduction. The sexes are separate in all starfishes. A pair of gonads lies in the coelom of each arm. Large numbers of eggs are shed directly into the sea, where they are fertilized and undergo early development At first the free-swimming larva is bilaterally symmetrical, becoming radially symmetrical only in the later stages of development The larvae ultimately develop heavy skeletons and settle to the bottom of the sea.



The starfish (From T. I. Storer and R L. Usinger, General Zoology, 3d ed., McGraw-Hill, New York, 1957)

Characteristics. Starfishes live on the bottom
n relatively shallow water. They are nearly all
predacious, their principal food being oysters and
other mollusks When attacked, they may volun-
tarily shed an arm, especially if it is injured This
is called autotomy. They are capable of regenera-
tion to a marked degree, even developing a new
animal if one arm and part of the disk is intact
One species will develop a new animal from only a
... regenerated arms

... hiuroidea
are similar to the starfishes except that the central
disk is quite small and the arms are typically long,
slender, and fragile. In the serpent stars, the arms
are unusually long and slender The basket stars
have arms which are repeatedly branched dicoto-
mously (forked) and terminate in tendril-like tips. [J D B]
See ECHINODERMATA

Stark effect

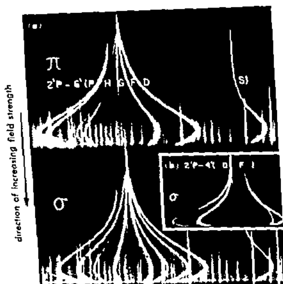
The effect of an electric field on spectrum lines The electric field may be externally applied, but in many cases it is an internal field caused by the presence of neighboring ions or atoms in a gas, liquid, or solid Discovered in 1913 by J Stark, the effect is most easily studied in the spectra of hydrogen and helium, by observing the light from the cathode dark space of an electric discharge. Because of the large potential drop across this region, the lines are split into several components. For observation perpendicular to the field, the light of these components is linearly polarized The splitting can be easily resolved with a spectrograph of moderate dispersion, amounting in the case of the H_γ line at 4340 A to a total spread of 32 A for a field of 104,000 volts/cm.

Linear Stark effect. This effect exhibits large, nearly symmetrical patterns. Examples are shown in the illustration, where the symbols π and σ refer to the two states of polarization (*see* ZEEMAN EFFECT). The interpretation of the linear Stark effect was one of the first successes of the quantum theory According to this theory, the effect of the electric field on the electron orbit is to split each energy level of the principal quantum number n into $2n-1$ equidistant levels, of separation proportional to the field strength Thus the higher members of a series show a larger number of components and greater over-all splittings The criterion for the occurrence of the linear Stark effect is that the splitting of the levels shall be large compared to the natural separation of levels of the same n but different L, where L is the quantum number of orbital angular momentum *See* QUANTUM NUMBERS

Quadratic Stark effect. This occurs in lines resulting from the lower energy states of many-electron atoms Here the large separation of states of different L results from the penetration of the valence electrons into the core of other electrons, with the result that the permanent dipole moment associated with hydrogenlike orbits no longer exists There is, however, a small induced dipole moment due to polarization of the atom. This moment is proportional to the electric field strength, and ... the energy change is proportional to the prod-
... strength the
... the
... s of
square of the field. Thus all ... the same sign and therefore are displaced to lower energies Each field-free level is also split, as a result of the space quantization of the angular momentum vector J, so that there are $J+1$ components if J is integral, or $J+\frac{1}{2}$ components if it is half-integral Because the lower levels are usually less displaced than the higher ones, the quadratic Stark effect ordinarily shows itself as a shift of the lines toward the red end of the spectrum, with an accompanying separation into several components.

The quadratic Stark effect is basic to the explanation of the formation of molecules from atoms, of dielectric constants, and of the broadening of spectral lines.

Intermolecular Stark effect. Produced by the action of the electric field from surrounding atoms or ions on the emitting atom, the intermolecular effect causes a shifting and broadening of spectrum lines. The molecules being in motion, these fields
... and also in time Hence
... but is
... c dis-
... large
ion density may cause very wide lines. The amount



Stark effect for helium lines (a) The 4144 and 4169 A lines at field strengths 0–85,000 volts/cm (b) The 4922 A line at 0–40,000 volts/cm In (a) note the disappearance of the $^1P-{}^1H$ line at a certain field strength, and the symmetry of the σ components (those with polarization perpendicular to the field), marked with dots In (b) note the crossing over of the two $^1P-{}^1F$ components. *(J. S. Foster and Curtis Foster, McGill University)*

greater use For example, starch may be converted with enzymes (amylases) to produce lower-viscosity types for sizing purposes. In the presence of acid, and at temperatures below the gelatinization point, starch may be converted in water suspension by breaking some of the molecules to obtain starches which give less viscous solutions. These thin boiling-type starches find many uses where increased concentration is desirable without more viscosity. These cooked starch pastes usually set on cooling with increased jellylike textures. The molecular size of starch may also be reduced with oxidizing agents. Carboxyl groups are introduced into the molecules in this way, and cooked starch pastes of these types are less jellylike in texture.

When starch is roasted in dry form, usually in the presence of a small amount of acid, a wide variety of starch products is obtained. The color usually darkens to shades of tan to brown, and the viscosity is reduced with increased roasting. This dextrinization process is a complex chemical reaction, and consists of a combination of hydrolysis, oxidation, and transglucosidation of the starch molecule, as well as repolymerization of some of the reaction products. The normally insoluble (in cold water) starch granule is converted with increasing dextrinization into smaller molecules which are soluble in cold water Some common names for dextrins are British gums, white dextrins, and canary or yellow dextrins. These products find many industrial uses Concentrated solutions of dextrins are sticky and have wide adhesive applications.

If the starch molecule is depolymerized further, usually by acid hydrolysis, a variety of sugars and dextrose or glucose is obtained. These in syrup or dry form find wide application in food products as sweeteners *See* CORN.

Carboxyl (COOH) groups are introduced into the starch molecule on oxidation. Other groupings or radicals, such as ethers or esters, can be introduced to produce different properties to the individual starch. Starch molecules also may be joined to form larger molecules. Thus, a wide variety of starch properties may be obtained to suit the [illegible] [T A.W.]

[illegible] *tarch and* [illegible] ch Assoc [illegible] Pulp Paper Ind., Monograph Ser., 17, 1957; R. W Kerr, *Chemistry and Industry of Starch*, 2d ed., 1950; National Starch Products Inc., *The Story of Starches*, 1953.

Starfish

A member of the class Asteroidea, phylum Echinodermata There are about 200 living species, all marine.

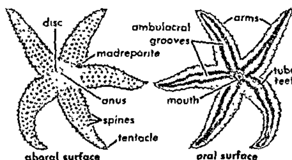
Economic importance. Starfishes are of some economic importance in areas where oysters are produced because they prey heavily upon oysters and can cause serious losses. Other than this they are of little importance. Some are used in zoology classes for study specimens. The early embryology is also studied as an example of holoblastic cleavage, and young stages are frequently utilized in various embryological experiments (*see* EMBRYOLOGY, EXPERIMENTAL; INVERTEBRATE EMBRYOLOGY). Dried specimens are frequently used as decorations.

Structure. Starfishes are radially symmetrical, with movable calcareous skeletons. Typically they are 5-armed, but some have a large, 5-sided disk without arms, whereas others may have 4–50 arms. Most starfishes are only a few inches across, but one has a spread up to 32 in.

Starfishes have a water vascular system, unique to the Echinodermata, which is used in locomotion, food handling, and respiration. In this system, water is used to fill a ring canal around the disk and branches which run down each arm. The branches connect to the ampullae of the numerous tube feet, which are small suction-cup structures. This hydraulic system enables the starfish to exert maximum force in such activities as opening oysters and pulling itself along with minimum muscular strength. The system opens on the dorsal surface of the disk through a pore, the madreporite.

In addition to the tube feet, respiration is accomplished by dermal papulae which are small saclike structures projecting through openings in the skeleton and leading to the coelom. The digestive system extends into the arms The mouth is ventral and connects directly to the centrally located stomach. Starfishes are covered with cilia on both surfaces, which probably aid in keeping their surfaces free of debris, plants, and animals. Most starfishes also have small pincerlike structures, the pedicillariae, on their surfaces, which aid in keeping them clean The circulatory system is reduced to a ring around the mouth and a small vessel running down each arm A similarly situated nerve ring with four nerves, which extend along each arm, comprises the nervous system.

Reproduction. The sexes are separate in all starfishes. A pair of gonads lies in the coelom of each arm Large numbers of eggs are shed directly into the sea, where they are fertilized and undergo early development At first the free-swimming larva is bilaterally symmetrical, becoming radially symmetrical only in the later stages of development. The larvae ultimately develop heavy skeletons and settle to the bottom of the sea.



The starfish (From T. I. Storer and R L. Usinger, General Zoology, 3d ed, McGraw-Hill, New York, 1957)

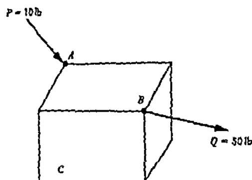


Fig 1 Graphic representation of forces.

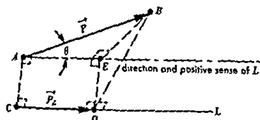


Fig 2 Directed line AB is a vector quantity

In Fig 2, the directed arrow $\overrightarrow{P}$ along line $\overline{AB}$ graphically represents any force whose line of action is parallel to the arrow whose sense of exertion is from A to B, and whose magnitude is the length AB, measured in specified units. An arrow so constructed is called a vector, it is designated P and, representing a force, it is called a force vector

External effects on a rigid body due to an imposed force may be a change in its motion state, a change in other forces acting on it, or generation of new forces. Theorems which pertain to these effects are as follows:

1 The principle of transmissibility of a force states that the external effects on a rigid body are the same for any position of the application point along the line of an imposed force In Fig 3, the external effects on rigid body C due to push force P acting at point A are the same as the effects caused by an equal pull force P applied at B

2 The principle of superposition states that the external effects on a rigid body are unchanged by imposing in pairs collinear forces of equal magnitude and opposite sense In Fig 3 the external effects on rigid body D are unaltered by applying either the pair Q Q of collinear and equal push forces or the pair R R of collinear and equal pull forces

The forces acting on a body in equilibrium are related Their dependence is commonly expressed in terms of the components and moments of the forces

Components of a force. Figure 2 shows the construction by which orthogonal force components are defined In the figure, $\overrightarrow{P}$ or $\overrightarrow{AB}$ is a force vector and L a directed line whose positive sense, arbitrarily chosen is toward its labeled end. Construction lines AC and BD are in planes (not shown) normal to L, and θ is the direction angle of $\overrightarrow{P}$ relative to L; it is a plane angle between the directions in the positive senses of $\overrightarrow{P}$ and L; further, $0 \leq \theta \leq 180°$.

The orthogonal vector component of force $\overrightarrow{P}$ on directed line L is a force of direction and magnitude given by $\overrightarrow{P_L}$ or $\overrightarrow{CD}$, where $\overrightarrow{P_L}$ is in the direction of L. Its magnitude is

$$P_L = \overline{CD} = \overline{AE} = |\overline{AB}\cos\theta| = P|\cos\theta|$$

The orthogonal scalar component of $\overrightarrow{P}$, on L, is $P_L = P\cos\theta$, where P_L is positive if $0° \leq \theta \leq 90°$ and negative, if $90° < \theta \leq 180°$. The absolute magnitude of P_L is designated $|P_L|$.

The rectangular components of a force are its components on mutually perpendicular lines.

In Fig. 4 $\overrightarrow{P_X}$ or $\overrightarrow{OA}$, $\overrightarrow{P_Y}$ or $\overrightarrow{OB}$, and $\overrightarrow{P_Z}$ or $\overrightarrow{OC}$ are the rectangular vector components of $\overrightarrow{P}$ or $\overrightarrow{OD}$ in the directions of lines (axes) X, Y, and Z, respectively.

The corresponding scalar components are $P_X = P\cos\theta_X$, $P_Y = P\cos\theta_Y$, and $P_Z = P\cos\theta_Z$.

From the geometry of Fig 4

$$P = +\sqrt{P_x^2 + P_y^2 + P_z^2}$$
$$\theta_x = \arccos(P_x/P)$$
$$\theta_y = \arccos(P_y/P)$$
$$\theta_z = \arccos(P_z/P)$$

Moment of a force. The moment of a force about a directed line is a signed number whose value

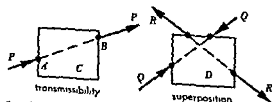


Fig 3 Illustrations of two theorems of statics.

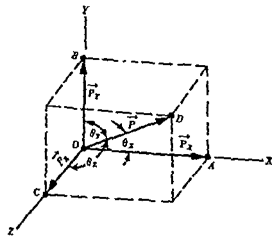


Fig 4 Rectangular components of a force vector.

of the broadening is found to run parallel to the sensitivity of the line to the Stark effect and thus is greatest for those lines susceptible to the linear effect.

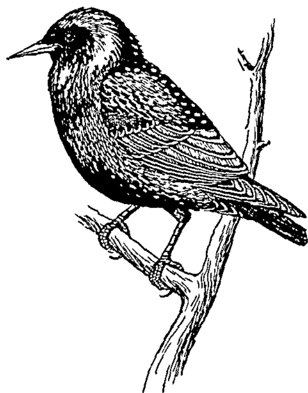
Inverse Stark effect. This is the effect as observed with absorption lines It has been detected, for example, by applying an electric field to potassium vapor, and measuring a small displacement of the absorption lines towards the red The displacements are found to be proportional to the square of the field strength, as in the quadratic Stark effect. In addition, certain lines of large n, which are normally forbidden by the selection rule for L, are observed to appear in the presence of the field. This type of transition is said to be produced by "forced dipole radiation." Such forbidden lines also may be obtained in emission In the illustration, the lines $2^1P - 6^1S$ and $2^1P - 4^1F$ are examples of forced dipole transitions. [F.A J.]

Bibliography: G. Herzberg, *Atomic Structure and Atomic Spectra*, reprint, 1944.

Starling

Any member of the Old World family Sturnidae, a family of perching birds. There are over 100 known species One of these, the European starling, *Sturnus vulgaris*, is widely distributed in the United States.

The first starlings were released in New York in 1890 They have since spread over almost all of the United States, although they are still uncommon in parts of the West. Starlings congregate in cities and roost on the larger public buildings, creating quite a nuisance. The starling is easily distinguished from other black or blackish birds by its short tail and yellow bill. *See* PASSERIFORMES [J.D B.]



The starling, *Sturnus vulgaris*; length to 8½ in (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Static

A hissing, crackling, or other sudden sharp noise that tends to interfere with the reception, utilization, or enjoyment of desired signals or sounds Perhaps the commonest form of static is that heard in ordinary broadcast receivers during electrical storms. Interference in radio receivers caused by improperly operating electric devices in the vicinity is sometimes also called static.

The crackling sounds heard when long-playing plastic phonograph records are played are also called static These sounds are caused by sudden deflection of the phonograph needle by dust particles, which are attracted to the grooves of the record by surface electric charges caused by friction on dry days. Static appears as momentary white specks in a television picture. *See* INTERFERENCE, ELECTRICAL; NOISE, ELECTRICAL. [J.MR]

Static electricity

The study of electric charges at rest and the fields they produce The fundamental fact of static electricity (or electrostatics) is that similarly electrified bodies repel, whereas oppositely electrified bodies attract, each other. Coulomb's law of force is the basic quantitative law of electrostatics *See* CHARGE, ELECTRIC, COULOMB'S LAW; ELECTROSTATICS. [R P.WI]

Statics

That branch of mechanics which treats of forces and the equilibrium of matter A body at rest on earth is considered to be in equilibrium

A body is acted on by many forces whose collective characteristics determine its motion. Mechanics deals with force, matter, and motion Statics treats of force and force systems abstracted from matter, and of forces which act on bodies in equilibrium

A problem of architects, engineers, and physicists is the identification and description of important forces which act on the parts of bodies at rest, such as buildings, dams, bridges, cranes, jacks, cables, and pressure vessels. Bodies like these, which essentially maintain their dimensions under force application, are idealized as rigid

Force as a vector. Force is a push or pull which matter exerts on other matter Force has magnitude and space direction Also, a force applied to a body acts at a certain point of the body, and along a certain line.

An arrow is used both to symbolize a force acting on a body and to represent, graphically, its properties of magnitude and direction. Figure 1 illustrates the symbolic usage. A push force, of magnitude 10 lb, acts along arrow P, and is applied at the point A on block C. Acting at B is 50-lb pull force Q.

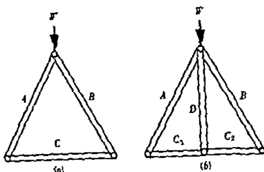


Fig 2 Statically determinate and indeterminate structures In (a) the forces in members A and B are independent of their strengths (for small deformations) In (b) the relative strength of the added member D will determine how much of the load W it supports.

can be found in both cases Better terminology would be rigidly determinable and indeterminable. In case 2b, if each member were rigid (infinitely strong), the forces in each could not be determined

Extended bodies have very many interconnections and so are statically indeterminate They may be solids or fluids and may behave more like surfaces than volumes, as in the case of soap films or stretched membranes For a discussion of fluids at rest *see* BUOYANCY, HYDROSTATICS For a discussion of small forces and deformations within a solid, *see* ELASTICITY

The equilibrium shapes of soap films depend on their special deformability There is no resistance to bending but a constant resistance to stretching in any direction because of *surface tension* A film attached to a *frame tends* to assume a shape with the smallest area, called a minimal surface These shapes are interesting both in mathematics and in architecture. A structure of such an equilibrium shape held in compression instead of tension would be in uniform compression throughout Certain building materials such as prestressed concrete are very strong in compression [B G]

Bibliography S Timoshenko and D. H Young, *Engineering Mechanics, Statics*, 1937

Stationary state

In quantum mechanics, an energy state for which the probability of any observation is independent of time that is stationary Because stationary states endure in time, they conveniently characterize physical systems. *For* example, at any instant, it is customary and useful to think that each atom in a discharge tube is in one of its stationary states, with most atoms in the ground state, and the radiation from the tube a measure of the populations of the excited states

A stationary state may be bound or unbound. bound states are localized, unbound states are not. Usually the potential energy of the particles comprising the system is zero at infinite separation, in which event the energies of bound states are negative, of unbound states, nonnegative For instance, in the physical system consisting of a proton and an electron, the bound states are the negative energy states of neutral atomic hydrogen, the unbound states are the states of the hydrogen ion H^+ in which the electron no longer is found mainly in the vicinity of the proton. Whether bound or unbound, stationary states are states of definite energy; that is, when several atoms are in the same stationary state, their measured energies always will be equal. This property need not hold for other observables; that is, the measured momentum (relative to the proton) of an electron in the ground state of hydrogen does not always have the same value *See* ENERGY LEVEL (QUANTUM MECHANICS); EXCITED STATE; GROUND STATE; QUANTUM MECHANICS; QUANTUM THEORY, NONRELATIVISTIC.

[E G]

Stationary wave

The boundary conditions within an enclosure are sometimes such that the amplitudes of waves traveling in opposite directions add in phase to form a third wave, the amplitude of which is stationary in time. This third wave is called a stationary wave. There is no intensity in a stationary wave, since no energy is absorbed at the reflections, and no net energy flow exists in the wave. For the case where net energy flow does exist, *see* STANDING WAVE.

Stationary waves can exist in one-, two-, or three-dimensional systems. The geometry and physical dimensions of the enclosure determine the actual stationary wave system. The simplest one-dimensional system results from sound waves being propagated between, and normal to, two parallel surfaces The frequencies f_n at which stationary waves may exist are related to the separation distance d between the parallel surfaces, and the speed of sound c, by the equation $f_n = nc/2d$, where n is an integer starting with unity. For example, if two reflecting surfaces are 10 ft apart, the lowest frequency at which a stationary wave would exist when the speed of sound is 1120 ft/sec is 56 cycles per second (cps). Higher frequency stationary waves exist at integral multiples of 56, such as 112, 168, and 224 cps

The frequency at which a stationary wave can exist is called a normal frequency of the enclosure. The conditions under which the stationary wave can exist are called normal modes or resonance conditions of the enclosure. Normal modes exist whenever the enclosure geometry permits the in-phase addition of oppositely directed waves. It is not even necessary that the walls of the enclosure be parallel. For example, in a rectangular parallelepiped, in addition to the three sets of modes which exist simply because of the reflections from the opposite parallel sides, additional modes exist because of waves traveling around the space-intercepting walls at certain prescribed angles of incidence. At these angles, waves will return upon themselves to establish a stationary wave The frequencies f_n of the normal modes for such an enclosure are given by the equation

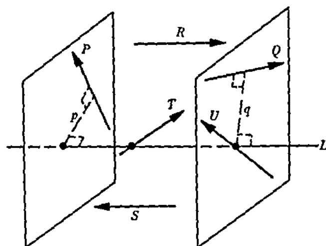


Fig. 5. Moments of forces about a line.

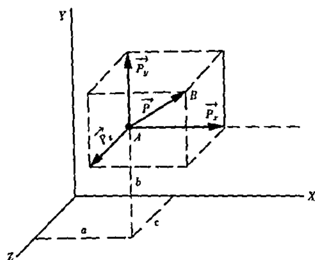


Fig 6. Moments about an axis in rectangular coordinates.

Statics in space

The study of equilibrium of a system under the action of impressed forces when the system, the forces, or both extend in three dimensions For a discussion of the case where all forces are in one plane, *see* STATICS; *see also* EQUILIBRIUM OF FORCES. The equilibrium conditions are almost the same for the plane case and the three-dimensional case, namely, for each part of the system separately, (1) the sum of forces acting in each direction is zero, and (2) the sum of torques about an axis in each direction is zero.

Planar forces are represented by vectors in a plane. Force vectors in space can have any direction. Torques about a point P, or more specifically, about axes through P can also be represented by vectors. The torque vector for a force whose vector does not pass through P is found as follows The line of action of the force and P together define a plane of action in which the torque is found just as in plane statics. This gives the length of the torque vector. Its direction is the axis perpendicular to the plane of action. The effect of independent torques about P is given by a resultant vector (Fig 1)

In terms of vectors, the conditions for static equilibrium are (for each part of the system): (1) the resultant of all forces is zero, and (2) the resultant of all torques about any point is zero.

Figure 2 illustrates cases where the forces in a structure may or may not depend on the relative strength of its parts. Case 2*a* is called statically determinate and case 2*b* statically indeterminate. The terms are actually misnomers, for the forces can be obtained by applying these two rules.

1. The moment of a force about a line parallel to the force is zero. In Fig 5 forces **S** and **R**, parallel to line L, have zero moments about L.

2 The moment of a force about a line normal to a plane containing the force is the product of the magnitude of the force and the least distance from the line to the line of the force Conventionally, the moment is positive if the force points counterclockwise about the line as viewed from the positive end of the line. The moment of **P** about line L in Fig. 5 is $M_L = +|P|p$, and of **Q** is $M_L = -|Q|q$

The moment of a force about a line through which the force passes is zero. For example, in Fig. 5 forces **T** and **U** have zero moment about L

Varignon's theorem is that the moment of a force about a directed line is the algebraic sum of the moments of its vector components acting at a common point on the line of the force

In Fig 6, the moment of **P** about the X axis is $M_X = P_Zb + (-P_Yc) = P_Zb - P_Yc$; also, $M_Y = P_Xc - P_Za$ and $M_Z = P_Ya - P_Xb$. Point A is on the line of **P**, and the components contain A on their lines of action. [N.S.F.]

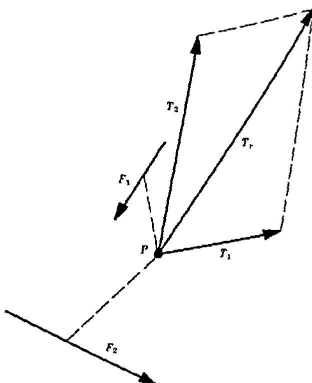


Fig. 1. Resultant of two torques

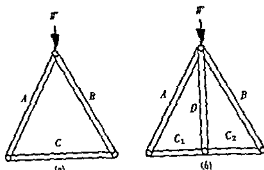


Fig 2 Statically determinate and indeterminate structures In (a) the forces in members A and B are independent of their strengths (for small deformations). In (b) the relative strength of the added member D will determine how much of the load W it supports

can be found in both cases Better terminology would be rigidly determinable and indeterminable. In case 2b, if each member were rigid (infinitely strong), the forces in each could not be determined.

Extended bodies have very many interconnections and so are statically indeterminate They may be solids or fluids and may behave more like surfaces than volumes, as in the case of soap films or stretched membranes For a discussion of fluids at rest *see* BUOYANCY, HYDROSTATICS For a discussion of small forces and deformations within a solid, *see* ELASTICITY

The equilibrium shapes of soap films depend on their special deformability There is no resistance to bending but a constant resistance to stretching in any direction because of surface tension A film attached to a frame tends to assume a shape with the smallest area, called a minimal surface. These shapes are interesting both in mathematics and in architecture. A structure of such an equilibrium shape held in compression instead of tension would be in uniform compression throughout Certain building materials such as prestressed concrete are very strong in compression [E.G.]

Bibliography S Timoshenko and D. H Young, *Engineering Mechanics, Statics*, 1937

Stationary state

In quantum mechanics an energy state for which the probability of any observation is independent of time that is stationary Because stationary states endure in time, they conveniently characterize physical systems For example at any instant, it is customary and useful to think that each atom in a discharge tube is in one of its stationary states, with most atoms in the ground state, and the radiation from the tube a measure of the populations of the excited states

A stationary state may be bound or unbound, bound states are localized, unbound states are not. Usually, the potential energy of the particles comprising the system is zero at infinite separation, in which event the energies of bound states are negative, of unbound states nonnegative For instance, in the physical system consisting of a proton and an electron, the bound states are the negative energy states of neutral atomic hydrogen, the unbound states are the states of the hydrogen ion H^+ in which the electron no longer is found mainly in the vicinity of the proton. Whether bound or unbound, stationary states are states of definite energy; that is, when several atoms are in the same stationary state, their measured energies always will be equal This property need not hold for other observables, that is, the measured momentum (relative to the proton) of an electron in the ground state of hydrogen does not always have the same value. *See* ENERGY LEVEL (QUANTUM MECHANICS); EXCITED STATE; GROUND STATE; QUANTUM MECHANICS; QUANTUM THEORY, NONRELATIVISTIC. [E.G.]

Stationary wave

The boundary conditions within an enclosure are sometimes such that the amplitudes of waves traveling in opposite directions add in phase to form a third wave, the amplitude of which is stationary in time. This third wave is called a stationary wave There is no intensity in a stationary wave, since no energy is absorbed at the reflections, and no net energy flow exists in the wave For the case where net energy flow does exist, *see* STANDING WAVE.

Stationary waves can exist in one-, two-, or three-dimensional systems. The geometry and physical dimensions of the enclosure determine the actual stationary wave system. The simplest one-dimensional system results from sound waves being propagated between, and normal to, two parallel surfaces. The frequencies f_n at which stationary waves may exist are related to the separation distance d between the parallel surfaces, and the speed of sound c, by the equation $f_n = nc/2d$, where n is an integer starting with unity. For example, if two reflecting surfaces are 10 ft apart, the lowest frequency at which a stationary wave would exist when the speed of sound is 1120 ft/sec is 56 cycles per second (cps) Higher frequency stationary waves exist at integral multiples of 56, such as 112, 168, and 224 cps.

The frequency at which a stationary wave can exist is called a normal frequency of the enclosure. The conditions under which the stationary wave can exist are called normal modes or resonance conditions of the enclosure Normal modes exist whenever the enclosure geometry permits the in-phase addition of oppositely directed waves It is not even necessary that the walls of the enclosure be parallel. For example, in a rectangular parallelepiped, in addition to the three sets of modes which exist simply because of the reflections from the opposite parallel sides, additional modes exist because of waves traveling around the space-intercepting walls at certain prescribed angles of incidence. At these angles, waves will return upon themselves to establish a stationary wave. The frequencies f_n of the normal modes for such an enclosure are given by the equation

$$f_n = \frac{c}{2}\left[\left(\frac{n_x}{l_x}\right)^2 + \left(\frac{n_y}{l_y}\right)^2 + \left(\frac{n_z}{l_z}\right)^2\right]^{1/2}$$

where n_x, n_y, n_z are integers from zero to infinity, and may be chosen independently of each other; l_x, l_y, l_z are the linear dimensions of the enclosure; and c is the speed of the wave.

Stationary waves can also exist in nonparallel walled enclosures such as cylinders, spheres, and ellipsoids. The normal modes are determined from the zeros of the solutions to the wave equation in each system of coordinates describing the geometry of the enclosure. These solutions generally involve transcendental functions, resulting in normal modes whose frequencies are not, in general, integrally related. *See* WAVE MOTION. [W.J.G.]

Bibliography: P. M. Morse, *Vibration and Sound*, 2d ed., 1948.

Statistical mechanics

That branch of physics which endeavors to explain the macroscopic properties of a system on the basis of the properties of the microscopic constituents of the system. Usually the number of constituents is very large. All the characteristics of the constituents and their interactions are presumed known; it is the task of statistical mechanics (often called statistical physics) to deduce from this information the behavior of the system as a whole.

SCOPE OF STATISTICAL MECHANICS

Elements of statistical mechanical methods are present in many widely separated areas in physics. For instance, in the classical Boltzmann problem one attempts to explain the thermodynamic behavior of gases on the basis of classical mechanics applied to the system of molecules (*see* BOLTZMANN STATISTICS). Historically this was the first systematic investigation of a statistical problem, and many of the procedures and methods originate from this investigation. It is important to realize that statistical mechanics gives more than just an explanation of already known phenomena. By using statistical methods, it often becomes possible to obtain expressions for empirically observed parameters, such as viscosity coefficients, heat conduction coefficients, and virial coefficients, in terms of the forces between molecules (*see* INTERMOLECULAR FORCES; KINETIC THEORY OF MATTER). This kind of result, a direct relation between an observed macroscopic entity and an intermolecular potential, constitutes one of the main achievements of statistical physics.

Statistical considerations also play a significant role in the description of the electric and magnetic properties of materials. In this case one would hope to deduce such characteristics as the dielectric constant, electrical conductivity, and magnetic permeability from the known properties of the atom. In this connection one should observe that Maxwell's equations are macroscopic equations, for matter in the bulk. To obtain the Maxwell equations from the Lorentz (electron) equations is typically a statistical question; new information which can be obtained from this approach is, for example, the temperature dependence of the dielectric constant.

If the problem of molecular structure is attacked by statistical methods, the contributions of internal rotation and vibration to thermodynamic properties, such as heat capacity and entropy, can be calculated for models of various proposed structures. Comparison with the known properties often permits the selection of the correct structure. The statistical description of the activated complex in reaction mechanisms has enabled chemists to formulate a theory of absolute reaction rates.

Perhaps the most dramatic examples of phenomena requiring statistical treatment are the cooperative phenomena or phase transitions. In these processes, such as the condensation of a gas, the transition from a paramagnetic to a ferromagnetic state, or the change from one crystallographic form to another, a sudden and marked change of the whole system takes place. The appropriate description of such processes is one of the most difficult but most interesting problems of statistical physics.

Statistical considerations of quite a different kind occur in the discussion of problems such as the diffusion of neutrons through matter. In this case, one knows the probability of the various events which affect the neutron, such as the capture probability and scattering cross section. The problem here is to describe the physical situation after a large number of these individual events. The procedures used in the solution of these problems are very similar to—and in some instances taken over from—kinetic considerations. Similar problems occur in the theory of cosmic-ray showers. Here the probabilities of the basic processes—pair production, annihilation, ionization, meson production—are all presumed known. The major task is the computation of the combined effect of a large number of these individual events. Special techniques of statistical physics, the use of distribution functions and transport-type equations, find extensive applications in these areas. *See* BOLTZMANN TRANSPORT EQUATION.

It happens in both low-energy and high-energy nuclear physics that a considerable amount of energy is suddenly liberated. An incident particle may be captured by a nucleus, or a high energy proton may collide with another proton. In either case there is a large number of ways (a large number of degrees of freedom) in which this energy may be utilized. In the nuclear case there are usually many decay and excitation modes, in the case of the proton-proton collision, there is enough energy available for the creation of a number of mesons. To survey the resulting processes one can again invoke statistical considerations. The statistical problem here is to calculate the probability that, given a total amount of energy, a particular process will occur. Of course the statistical method cannot help in the investigation of the actual mechanisms of these processes. However, the sta-

tistical factors must be considered in the interpretation of experiments See SCATTERING EXPERIMENTS, NUCLEAR.

Of considerable importance in statistical physics are the random processes, also called stochastic processes or sometimes fluctuation phenomena. The Brownian motion, the motion of a particle moving in an irregular manner under the influence of molecular bombardment, affords a typical example (see BROWNIAN MOVEMENT, STOCHASTIC PROCESS) The process may be described in terms of a fluctuating force acting on a particle, perhaps in addition to a systematic force. The interest in this problem centers around a calculation of the position and velocity of a particle after it has experienced many collisions or after the fluctuating force has acted for a long time. The stochastic processes are in a sense intermediate between purely statistical processes, where the existence of fluctuations may safely be neglected, and the purely atomistic phenomena, where each particle requires its individual description All statistical considerations involve, directly or indirectly, ideas from the theory of probability, of widely different levels of sophistication The use of probability notions is, in fact the distinguishing feature of all statistical considerations See PROBABILITY, PROBABILITY IN PHYSICS, STATISTICS

METHODS OF STATISTICAL MECHANICS

Phase space. Consider a system of N particles, each of mass m, contained in a volume V. Call the positions of the particles $x_1, y_1, z_1, \ldots, x_N, y_N, z_N$, their cartesian velocities $v_{x_1}, \ldots, v_{z_N}$, and their momenta $p_{x_1}, \ldots, p_{z_N}$ The simplest statistical description concentrates on a discussion of the *distribution* function $f(x,y,z,v_x,v_y,v_z,t)$ The quantity $f(x,y,z,v_x,v_y,v_z,t)(dx\,dy\,dz\,dv_x\,dv_y\,dv_z)$ gives the (probable) number of particles of the system in those positional and velocity ranges where x lies between x and $x + dx$, v_x between v_x and $v_x + dv_x$, *etc* These ranges are finite The 6-dimensional space defined by x, y, z, v_x, v_y, v_z is called the μ space and the behavior of the gas is represented geometrically by the *motion* of N points in the μ space It was shown by L. Boltzmann that in the course of time f approaches an equilibrium distribution f^0:

$$\begin{aligned} &f^0(\mathbf{x}, \mathbf{v})\, d^3x\, d^3v \\ &\quad = A \exp\left[-(\tfrac{1}{2})\beta m v^2 - \beta U(x,y,z)\right] d^3x\, d^3v \qquad (1) \end{aligned}$$

Here $d^3x = dx\,dy\,dz$, A and β are parameters, and $U(x,y,z)$ is the potential energy at the point x, y, z Boltzmann also could interpret Eq (1) as the most probable distribution

Actually important as the use of the distribution function is in practice, there *are in* principle serious limitations and difficulties associated with it The limitations refer to the fact that the description by means of a single distribution function is correct only for noninteracting particles, also, the neglect of triple and higher collisions restricts the applicability to very dilute systems. The indiscriminate use of the ideas of Boltzmann leads to paradoxical results.

In the further study of these questions, the *phase space* of a dynamical system plays an important role Consider the $6N$-dimensional euclidean space, whose $6N$ axes are

$$x_1, y_1, z_1, \ldots, x_N, y_N, z_N;\ p_{x_1}, \ldots, p_{z_N} \qquad (2)$$

The state of the system at a given time is completely specified by these $6N$ numbers. These numbers define a point (the representative point, or phase point) in the $6N$-dimensional phase space. In the course of time the variables $x_1, \ldots, p_{z_N}$ change, hence the phase point moves. If the system is conservative, Eq. (3) defines, *for the energy* E, a $(6N-1)$-dimensional surface in the phase space called the *energy surface*.

$$\sum_i \frac{p_i^2}{2m_i} + U(x_1, \ldots, z_N) = E \qquad (3)$$

For a conservative system, the phase point in the course of time wanders over the energy surface. Intuitive as this geometrical representation of an involved mechanical system might seem, it is well to keep in mind that the actual *trajectory of the* phase point is extremely complicated Also, the use of the phase space does not by itself imply any statistical considerations. A knowledge of the trajectory is precisely equivalent to a *solution of the* complete mechanical problem, for one would know each p and x as a function of time. It should be clear, however, that many different microscopic states all correspond to the *same macroscopic situation*

Observations made on a system always require a finite time; during this time the microscopic details of the system will *generally change* considerably as the phase point moves The result of a measurement of a quantity Q will therefore yield the time average:

$$\bar{Q} = \frac{1}{t}\int_0^t Q\, dt \qquad (4)$$

The integral *is* along the trajectory in phase space; Q depends on the variables $x_1, \ldots, p_{z_N}$ and t. To evaluate the *integral one* must know the trajectory, which, as already mentioned, *requires* the solution of *the complete* mechanical problem

For many *years one of the* major *problems* in statistical mechanics was *precisely* to recast the expression for a time average in a more tractable form. Generally one *attempted* to replace time averages by phase space averages, by trying *to* show that the time spent *in a region* was proportional to the size *of the region* The legitimacy of this procedure was studied *extensively* by mathematicians, but although *these* studies opened up significant areas in mathematics, they did *not* really help in the elucidation of *problems* of physical interest.

To illustrate the subtlety *of the* problems involved, attention *may* be called to the so-called

$$f_n = \frac{c}{2}\left[\left(\frac{n_x}{l_x}\right)^2 + \left(\frac{n_y}{l_y}\right)^2 + \left(\frac{n_z}{l_z}\right)^2\right]^{1/2}$$

where n_x, n_y, n_z are integers from zero to infinity, and may be chosen independently of each other; l_x, l_y, l_z are the linear dimensions of the enclosure; and c is the speed of the wave.

Stationary waves can also exist in nonparallel walled enclosures such as cylinders, spheres, and ellipsoids. The normal modes are determined from the zeros of the solutions to the wave equation in each system of coordinates describing the geometry of the enclosure. These solutions generally involve transcendental functions, resulting in normal modes whose frequencies are not, in general, integrally related. *See* WAVE MOTION. [W.J.G.]

Bibliography: P. M. Morse, *Vibration and Sound*, 2d ed., 1948.

Statistical mechanics

That branch of physics which endeavors to explain the macroscopic properties of a system on the basis of the properties of the microscopic constituents of the system. Usually the number of constituents is very large. All the characteristics of the constituents and their interactions are presumed known; it is the task of statistical mechanics (often called statistical physics) to deduce from this information the behavior of the system as a whole.

SCOPE OF STATISTICAL MECHANICS

Elements of statistical mechanical methods are present in many widely separated areas in physics. For instance, in the classical Boltzmann problem one attempts to explain the thermodynamic behavior of gases on the basis of classical mechanics applied to the system of molecules (*see* BOLTZMANN STATISTICS). Historically this was the first systematic investigation of a statistical problem, and many of the procedures and methods originate from this investigation. It is important to realize that statistical mechanics gives more than just an explanation of already known phenomena. By using statistical methods, it often becomes possible to obtain expressions for empirically observed parameters, such as viscosity coefficients, heat conduction coefficients, and virial coefficients, in terms of the forces between molecules (*see* INTERMOLECULAR FORCES; KINETIC THEORY OF MATTER). This kind of result, a direct relation between an observed macroscopic entity and an intermolecular potential, constitutes one of the main achievements of statistical physics.

Statistical considerations also play a significant role in the description of the electric and magnetic properties of materials. In this case one would hope to deduce such characteristics as the dielectric constant, electrical conductivity, and magnetic permeability from the known properties of the atom. In this connection one should observe that Maxwell's equations are macroscopic equations, for matter in the bulk. To obtain the Maxwell equations from the Lorentz (electron) equations is typically a statistical question; new information which can be obtained from this approach is, for example, the temperature dependence of the dielectric constant.

If the problem of molecular structure is attacked by statistical methods, the contributions of internal rotation and vibration to thermodynamic properties, such as heat capacity and entropy, can be calculated for models of various proposed structures. Comparison with the known properties often permits the selection of the correct structure. The statistical description of the activated complex in reaction mechanisms has enabled chemists to formulate a theory of absolute reaction rates.

Perhaps the most dramatic examples of phenomena requiring statistical treatment are the cooperative phenomena or phase transitions. In these processes, such as the condensation of a gas, the transition from a paramagnetic to a ferromagnetic state, or the change from one crystallographic form to another, a sudden and marked change of the whole system takes place. The appropriate description of such processes is one of the most difficult but most interesting problems of statistical physics.

Statistical considerations of quite a different kind occur in the discussion of problems such as the diffusion of neutrons through matter. In this case, one knows the probability of the various events which affect the neutron, such as the capture probability and scattering cross section. The problem here is to describe the physical situation after a large number of these individual events. The procedures used in the solution of these problems are very similar to—and in some instances taken over from—kinetic considerations. Similar problems occur in the theory of cosmic-ray showers. Here the probabilities of the basic processes—pair production, annihilation, ionization, meson production—are all presumed known. The major task is the computation of the combined effect of a large number of these individual events. Special techniques of statistical physics, the use of distribution functions and transport-type equations, find extensive applications in these areas. *See* BOLTZMANN TRANSPORT EQUATION.

It happens in both low-energy and high-energy nuclear physics that a considerable amount of energy is suddenly liberated. An incident particle may be captured by a nucleus, or a high energy proton may collide with another proton. In either case there is a large number of ways (a large number of degrees of freedom) in which this energy may be utilized. In the nuclear case there are usually many decay and excitation modes; in the case of the proton-proton collision, there is enough energy available for the creation of a number of mesons. To survey the resulting processes one can again invoke statistical considerations. The statistical problem here is to calculate the probability that, given a total amount of energy, a particular process will occur. Of course the statistical method cannot help in the investigation of the actual mechanisms of these processes. However, the sta-

ble for a thermodynamic equilibrium state must be described by a density function which is independent of the time, since all the macroscopic averages which are to be computed as ensemble averages are *time-independent*. Thus one has an equilibrium ensemble when $\partial\rho/\partial t = 0$. In that case it follows from Liouville's theorem that

$$\sum_{i=1}^{3N}\left(\frac{\partial\rho}{\partial x_i}\frac{dx_i}{dt} + \frac{\partial\rho}{\partial p_i}\frac{dp_i}{dt}\right) = 0 \qquad (10)$$

It is easy to see that if ρ is a function of a quantity α (which is a function of x and p and which is a constant of the motion, such as the energy, so that $\partial\alpha/\partial t = 0$), then any $\rho(\alpha)$ satisfies Eq. (10). Hence any density function ρ which is a function of x and p through the dependence of the energy E on p and x satisfies the Liouville equation. The functional form of $\rho(E)$ is left completely unspecified. If one deals with an isolated system, where the energy is specified within a well-defined range, the most obvious example to use would be the so-called microcanonical ensemble defined by

$$\rho(p,x) = c \qquad (11a)$$

where c is a constant, for the energy E between E_0 and $E_0 + \Delta E$, for other energies

$$\rho(p,x) = 0 \qquad (11b)$$

By using Eq. (6), one may calculate any microcanonical average. The calculations, which involve integrations over volumes bounded by two energy *surfaces* are not trivial. Still, one may obtain in this way many of the results of classical Boltzmann statistics.

For applications and for the *interpretation* of thermodynamics, the canonical ensemble is much more preferable. In this case one describes a system which is not isolated but which is in thermal contact with a heat reservoir. By describing the complete system (*system plus reservoir*) by a microcanonical *ensemble* it may be shown that the system itself may be represented by

$$\rho(x,p) = N_e \exp\left[\frac{\psi - E(x,p)}{\theta}\right] \qquad (12)$$

Here N_e is the total number of *ensemble* systems. $E(x,p)$ is the mechanical energy of the system, and ψ and θ are parameters independent of x and p *characterizing* the canonical ensemble. From the fact that

$$\int\rho(x,p)\,d\Gamma = N_e \qquad (12a)$$

one *deduces* that the two parameters ψ and θ are not independent, but that

$$e^{-\psi/\theta} = \int d\Gamma \exp\left[-\frac{E(x,p)}{\theta}\right] = Z \qquad (13)$$

The quantity Z is usually called the partition function. It is customary to define Z for a system of N identical particles with a different multiplicative constant

$$Z' = \frac{1}{N!h^{3N}}\int \exp\left[-\frac{E(p,x)}{\theta}\right] d\Gamma \qquad (14)$$

where h is the Planck constant. It is now very important that the parameters θ and ψ be identified with definite thermodynamic functions. In this connection, consider some of the thermodynamic relations. Call the entropy η, the thermodynamic energy (internal energy) ϵ, and the free energy (Holmholtz free energy) ψ'. Then some examples of thermodynamic relations are, for gases, described by pressure, volume, and temperature (P, V, T)

$$\psi' = \epsilon - T\eta \qquad (15a)$$

$$d\psi' = -\eta\,dT - P\,dV \qquad (15b)$$

$$\frac{\partial\psi'}{\partial T} = -\eta \qquad \frac{\partial\psi'}{\partial V} = -P \qquad (15c)$$

$$\epsilon = \psi' - T(\partial\psi'/\partial T) \qquad (15d)$$

(See THERMODYNAMIC PRINCIPLES.) From Eqs. (12), (6), and (13) one may calculate the average energy $\bar{E}$:

$$\bar{E} = \theta^2\frac{\partial}{\partial\theta}\log Z \qquad (16)$$

For an ideal gas, in which the energy of a molecule is $p_i^2/2m$,

$$Z = \int \cdots \int dx_1 \cdots dp_{3N} \exp\left(-\frac{1}{2m\theta}\sum_{i=1}^{N} p_i^2\right)$$
$$= V^N(2\pi m\theta)^{3N/2} \qquad (17)$$

(The evaluation of the integrals is elementary; the space integrations merely contribute V^N, and the momentum integrals are Gaussian.) One then obtains

$$\bar{E} = \frac{3N}{2}\theta \qquad (18)$$

as the energy of an ideal gas, having N molecules in a vessel of any volume V. Now one knows by experiment that the thermodynamic energy of 1 mole of an ideal gas is

$$\epsilon = \tfrac{3}{2}NkT = \tfrac{3}{2}RT \qquad (18a)$$

Here R is the ideal gas constant, T is the absolute temperature, N is Avogadro's number, and k is the Boltzmann constant. Comparison of Eqs. (18) and (18a) leads to the identification of $\theta = kT$. It is also easy to deduce the Maxwell-Boltzmann distribution. Applying Eq. (7), one finds for the probability that molecule 1 will have a given range of positions and momenta

$$P(p_1,x_1)\,d\,p_1\,d^3x_1$$
$$= \frac{\int \cdots \int d^3x_2\,d^3p_2\,d^3x_2 \cdots d^3p_N\,e^{-E/\theta}}{\int \cdots \int d^3x_1 \cdots d^3p_N\,e^{-E/\theta}}$$
$$= \frac{1}{V}e^{-p^2/(2m\theta)}(2m\theta)^{-3/2}\,d\,p_1\,d^3x_1 \qquad (19)$$

ergodic theorem of G. D. Birkhoff, which asserts that the limit $\lim_{t \to \infty} \bar{Q}_t$ exists for almost all trajectories, although this limit varies discontinuously from trajectory to trajectory. The Poincaré recurrence theorem, which states that a system within a finite time will return as closely as one pleases to its initial state, is of special interest in statistical mechanics. Apart from providing the mathematical framework for these theorems, the phase space provides also the appropriate framework for the formulation of ensemble theory, which is the basis of modern statistical mechanics.

Ensembles; Liouville's theorem. J. Willard Gibbs first suggested that instead of calculating a time average for a single dynamical system, one should instead consider a collection of systems, all similar to the original one. Such an ensemble of systems is to be constructed in harmony with the available knowledge of the single system, and may be represented by an assembly of points in the phase space, each point representing a single system. If, for example, one knows the energy of a system precisely, but nothing else, the appropriate representative example would be a uniform distribution of ensemble points over the energy surface, and no ensemble points elsewhere. An ensemble is characterized by a density function $\rho(x_1, \ldots, z_N, p_{x_1}, \ldots, p_{z_N}, t) \equiv \rho(x,p,t)$. The significance of this function is that the number of ensemble systems dN_e contained in the volume element $dx_1 \cdots dz_N$, $dp_x \cdots dp_{z_N}$ of the phase space (this volume element will be called $d\Gamma$), at time t is

$$\rho(x,p,t)\; d\Gamma = dN_e \qquad (5)$$

The ensemble average of any quantity Q is given by

$$\bar{Q}_{ens} = \frac{\int Q\rho\, d\Gamma}{\int \rho\, d\Gamma} \qquad (6)$$

The basic idea is now to replace the time average of an individual system by the ensemble average, at a fixed time, of the representative ensemble. Stated formally, one identifies $\bar{Q}_t$ defined by Eq. (4) in which no statistics is involved, with $\bar{Q}_{ens}$ defined by Eq (6), in which probability assumptions are explicitly made. Another form of this same connection between the behavior of the individual system and the ensemble is that the probability that the individual system at time t will be in a region R of the phase space is given by the fraction of ensemble systems contained in R at time t:

$$P(R,t) = \frac{\int_R \rho\, d\Gamma}{\int \rho\, d\Gamma} \qquad (7)$$

It is clear, on the basis of these relations, that the complete statistical behavior of a system is known once its representative ensemble has been obtained. In the construction of such ensembles, one always assumes that accessible parts of the phase space should be weighted equally. There is one other general requirement the density function must satisfy. Inasmuch as the number of ensemble members remains constant (no ensemble members are created or destroyed in the course of time), there is a continuity equation for the density function ρ which simply states that the change in the number of systems in a volume of phase space per unit time equals the difference of the number of systems flowing in and flowing out of that volume

$$\frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N} \left[\frac{\partial}{\partial x_i}\left(\rho \frac{dx_i}{dt}\right) + \frac{\partial}{\partial p_i}\left(\rho \frac{dp_i}{dt}\right)\right] = 0 \quad (8)$$

If one now expresses the time derivatives dx_i/dt and dp_i/dt through the Hamiltonian equations (see HAMILTON'S EQUATIONS OF MOTION), one obtains

$$\frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N} \left(\frac{\partial \rho}{\partial x_i}\frac{dx_i}{dt} + \frac{\partial \rho}{\partial p_i}\frac{dp_i}{dt}\right) \equiv \frac{d\rho}{dt} = 0 \qquad (9)$$

Here $d\rho/dt$ is the substantial, or total derivative. Equation (9), called the Liouville theorem, asserts that the time rate of change of ρ is zero along a streamline in phase space. One can also deduce from Eq. (9) that the ensemble systems move in the course of time in a volume-preserving fashion through the phase space This means that if at some initial time one considers a set of ensemble points which occupy a volume V_0, then in the course of time each representative point will move in a complicated fashion At time t one can consider the volume made up by the ensemble points which initially were in V_0. This volume V_t equals V_0. The shape of the volume changes, but the actual volume is constant. In the attempts which have been made to justify the postulate that $\bar{Q}_t = \bar{Q}_{ens}$, this volume-preserving property plays a dominant role. The proof of the Poincaré recurrence theorem also depends crucially on this property Yet in spite of the general significance of the Liouville theorem. Eq. (9) generally does not help in the task of setting up an appropriate representative ensemble. (See, however, the subsequent discussion on non equilibrium theory of liquids) The choice of the ensemble is usually determined by physical considerations and by the already mentioned postulate of equal a priori probabilities in phase space. Perhaps most important is the a posteriori justification of the statistical procedures, obtained through a comparison with the experimental facts. It is important that inasmuch as the methods are statistical one can compute not only average values but also fluctuations around these average values It turns out that usually these fluctuations are extremely small; however in those instances in which they are appreciable, they can be compared with experiments, providing additional support for the statistical procedures Hence despite the fact that no completely rigorous mathematical justification of the ensemble methods exists, these methods are physically so plausible and lead to such remarkable agreement with experiments for a variety of systems that they reasonably must be considered as a well-established part of statistical physics.

Relation to thermodynamics. It is certainly reasonable to assume that the appropriate ensem-

energy or thermodynamic potential; $\xi = \psi + PV$. Formally, these results are

$$P = -\left(\frac{\partial\Omega}{\partial V}\right)_{\mu, T} \tag{30a}$$

$$\eta = -\left(\frac{\partial\Omega}{\partial T}\right)_{V,\mu} \tag{30b}$$

$$\bar{N}_{gr} = -\left(\frac{\partial\Omega}{\partial\mu}\right)_{V, T} \tag{30c}$$

From a knowledge of the grand partition function as a function of V, T, and μ, all thermodynamic functions follow. For an ideal gas, for example, Eq. (28a) in conjunction with Eq. (17) yields

$$Z_{gr} = \sum_{N=1}^{\infty} \frac{1}{N!}\left(\frac{2\pi m\theta}{h^2}\right)^{(3/2)N} V^N e^{\mu N/\theta}$$
$$= \exp\left(e^{\mu/\theta}\frac{V}{\lambda^3}\right) \tag{31}$$

Here λ is the thermal de Broglie wavelength:

$$\lambda = \frac{h}{\sqrt{2\pi mkT}} \tag{31a}$$

From Eqs. (31) and (29) one sees that

$$\Omega = -\theta e^{\mu/\theta}\frac{V}{\lambda^3} \tag{32}$$

Application of Eq. (30) yields the usual results for ideal gases, such as $PV = \bar{N}kT$; however, the grand average of $\bar{N}$ enters the relations now rather than just N. From the definitions (27) and (28a) one may show that the fractional fluctuation in the number N is given by

$$\left(\frac{\overline{N_{gr}^2} - (\bar{N}_{gr})^2}{\bar{N}_{gr}^2}\right)^{1/2} = \frac{1}{\sqrt{\bar{N}_{gr}}} \tag{33}$$

Thus for gases the fluctuations in number are negligibly small, showing that the number of particles in the grand canonical ensemble is sharply peaked around the average value. One therefore expects that the physical results deduced from the canonical and grand-canonical ensemble will be the same; the calculations, however, are frequently (especially in quantum problems) simpler in the grand-canonical than in the canonical scheme.

Applications. Two of the newest and most important applications of statistical physics involve the [illegible] importance of relations such as Eq. (13), which connects the thermodynamic free energy to the partition function, lies in the general validity of these connections. Specifically, when dealing with a gas of interacting molecules, where the energy is given by

$$E(x_1, \ldots, x_N, p_{x_1}, \ldots, p_{x_N}) = \sum_{i=1}^{N} \frac{p_{x_i}^2 + p_{y_i}^2 + p_{z_i}^2}{2m} + U(x_1, \ldots, z_N)$$

with the potential U a known function, the relation is still valid. In that case, one may obtain the partition function directly in terms of U, by performing the integrations over the momenta:

$$Z_N = \frac{1}{N!}\frac{1}{\lambda^{3N}}\int \cdots \int d^3x_1 \cdots d^3x_N \exp\left[-\frac{U(x_1, \ldots, z_N)}{kT}\right] \tag{34}$$

Since from Z_N all thermodynamic relations follow, Eq. (34) is the appropriate starting point for all studies of nonideal gases. The canonical ensemble has been used in obtaining Z_N, which depends on V through the integration limits, on T through λ (see 31a), and on N through the number of integrations. Since Z_N gives the free energy immediately, one can get the equation of state (the relation between P, V, and T), by computing $P = -\partial\psi/\partial V$. The experimental results on the equations of state may be expressed in terms of the so-called virial expansion:

$$P = \frac{RT}{V}\left(1 + \frac{B(T)}{V} + \frac{C(T)}{V^2} + \cdots\right) \tag{35}$$

Here $B(T)$ and $C(T)$ are the second and third virial coefficients; they are determined experimentally. It is clear that (35) is a development starting from the ideal gas equation. A first task for statistical mechanics would be to deduce the form (35), as well as an explicit expression for $B(T)$, etc.

Let $$\exp\left[-\frac{U(x_1, \ldots, z_N)}{kT}\right] = W(1,2, \ldots, N) \tag{36}$$

Assume now that the potential $U(x_1, \ldots, z_N)$ is such that it is possible to make an unambiguous distinction between interacting and noninteracting particles. Any potential which vanishes for a separation larger than a critical amount has this property, and any short-range potential approximates it. Call a separated configuration one in which none of the molecules of one group interacts with any of another. If two such groups are called α and β, one has, for such a configuration,

$$U = U(\alpha) + U(\beta) \tag{36a}$$

hence, by (36),

$$W = W(\alpha)W(\beta) \tag{36b}$$

Consider now a set of functions S defined by

$$W(1) = S(1) = 1$$
$$W(1,2) = S(1,2) + S(1)S(2)$$
$$W(1,2,3) = S(1,2,3) + S(1)S(2,3) + S(2)S(1,3) + S(3)S(1,2) + S(1)S(2)S(3)$$
$$W(1, \ldots, N) = S(1, \ldots, N) + S(1)S(2, \ldots, N) + S(1,2)S(3, \ldots, N) + \cdots + S(1) \cdots S(N) \tag{36c}$$

From Eqs. (36b) and (36c), one now proves the important property: S functions vanish for a sepa-

The canceled symbols in the numerator mean that one does not integrate over the coordinates and momenta of the molecule 1. The result given by Eq. (19) is again true for an ideal gas. The identification of $\theta = kT$, however, is made in general. By differentiating Eq. (13) with respect to θ, one may establish the general connection (using $\theta = kT$)

$$\bar{E} = \psi - T\frac{\partial\psi}{T} = \epsilon \qquad \text{[by Eq. (18)]} \tag{20}$$

When this general relation, deduced from the canonical ensemble, is compared with the general thermodynamic relation (15d), one sees indeed that the parameter ψ and the free energy ψ' play identical roles. Again, for an ideal gas one can compute ψ, and it is the same as the ideal gas free energy. Hence one identifies ψ and ψ' in general. Equation (13) relates the thermodynamic free energy to an integral containing the mechanics of the microscopic problem. It is still interesting to observe that a quantity

$$\eta' = \frac{\int\rho \log \rho \, d\Gamma}{\int\rho \, d\Gamma} \tag{21}$$

plays in all instances the role of the entropy. In fact, one may show that $\psi = \epsilon - T\eta'$, in harmony with (15a). One can follow these thermodynamic analogies through in all detail to see that the canonical ensemble combined with the relevant definitions does indeed reproduce the formal aspects of thermodynamics. The average energy $\bar{E}$ for a canonical ensemble has already been calculated. It is important to know just how often appreciable deviations from this average energy can be expected. For this, one needs to calculate the fractional fluctuations in energy given by

$$\frac{\overline{E^2} - (\bar{E})^2}{(\bar{E})^2}$$

The definition of $\overline{E^2}$ follows again from Eq. (6):

$$\overline{E^2} = \frac{\int E^2 e^{-E/\theta} \, d\Gamma}{\int e^{-E/\theta} \, d\Gamma} \tag{22}$$

From Eq. (22) one may directly express $\overline{E^2}$ in terms of Z, the partition function. For an ideal gas one can then obtain, using Eq. (17), an explicit expression for the fluctuations in energy:

$$\left(\frac{\overline{E^2} - (\bar{E})^2}{(\bar{E})^2}\right)^{1/2} = \sqrt{\frac{2}{3N}} \tag{23}$$

Since one deals with systems where the number of particles N is about 10^{22}, the chance of observing a sizable deviation from the average energy is extremely small. Using these same methods, defining in addition the specific heat $C_v = \partial\bar{E}/\partial T$, one may show generally that the fractional fluctuation in energy is given by

$$\left(\frac{\overline{E^2} - (\bar{E})^2}{(\bar{E})^2}\right)^{1/2} = \left(\frac{kT^2C_v}{(\bar{E})^2}\right)^{1/2} \tag{24}$$

Even though the energy fluctuations are negligible for an ideal gas, this is not always so for other systems. For solids at low temperatures, the fluctuations may be appreciable. When phase transitions (of the first order) take place, C_v becomes infinite, indicating via Eq. (24) that the fluctuations become very large. It is clear that the main problem of the applications is reduced to the calculation of the partition function Z, all thermodynamic entities follow from Eqs. (16) and (13).

There is yet another ensemble which is extremely useful and which is particularly suitable for quantum mechanical applications. Much work in statistical mechanics is now based on the use of this so-called grand-canonical ensemble. In the grand ensemble one has a collection of systems; the number of particles in each system is no longer the same, but varies from system to system. The density function $\rho(N,p,x)\; d\Gamma_N$ gives the probability that there will be in the ensemble a system having N particles, and that this system, in its $6N$-dimensional phase space Γ_N, will be in the region of phase space $d\Gamma_N$. The function ρ is given by

$$\rho(N,p,x) = \exp\left(\frac{\Omega + \mu N - E(p,x)}{\theta}\right) \tag{25}$$

Here θ, Ω, and μ are the parameters characterizing the ensemble, just as ψ and θ characterize the canonical ensemble. It is again true that these parameters are directly related to thermodynamic state functions. The detailed argument follows the identical pattern indicated in the discussion of the canonical ensemble. The normalization condition is now

$$\sum_N \int d\Gamma_N \, \rho(N,p,x) = \sum_N \int d\Gamma_N \exp\left(\frac{\Omega + \mu N - E(p,x)}{\theta}\right) = 1 \tag{26}$$

The grand canonical average of any quantity Q is

$$\bar{Q}_{gr} = \sum_N \int d\Gamma_N \, \rho(N,p,x) Q_N(p,x) \tag{27}$$

and the grand partition function is

$$Z_{gr} = \sum_N e^{\mu N/\theta} \int d\Gamma_N \, e^{-(E/\theta)} \tag{28}$$

Again one customarily uses for a system consisting of N identical particles

$$Z_{gr} = \sum_N \frac{1}{N!}\frac{1}{h^{3N}} e^{\mu N/\theta} \int d\Gamma_N \, e^{-(E/\theta)} \tag{28a}$$

From Eq. (26) one sees immediately that

$$Z_{gr} = e^{-(\Omega/\theta)} \tag{29}$$

From Ω, sometimes called the grand potential, all other thermodynamic functions may be computed. The parameter μ is the chemical potential. It is defined by $\mu = (\partial\xi/\partial N)_{P,T}$ where ξ is the Gibbs free

energy or thermodynamic potential; $\xi = \psi + PV$. Formally, these results are

$$P = -\left(\frac{\partial\Omega}{\partial V}\right)_{\mu,T} \qquad (30a)$$

$$\eta = -\left(\frac{\partial\Omega}{\partial T}\right)_{V,\mu} \qquad (30b)$$

$$\bar{N}_{gr} = -\left(\frac{\partial\Omega}{\partial\mu}\right)_{V,T} \qquad (30c)$$

From a knowledge of the grand partition function as a function of V, T, and μ, all thermodynamic functions follow. For an ideal gas, for example, Eq (28a) in conjunction with Eq (17) yields

$$Z_{gr} = \sum_{N=0}^{\infty} \frac{1}{N!}\left(\frac{2\pi m\theta}{h^2}\right)^{(3/2)N} V^N e^{\mu N/\theta}$$
$$= \exp\left(e^{\mu/\theta}\frac{V}{\lambda^3}\right) \qquad (31)$$

Here λ is the thermal de Broglie wavelength:

$$\lambda = \frac{h}{\sqrt{2\pi mkT}} \qquad (31a)$$

From Eqs (31) and (29) one sees that

$$\Omega = -\theta e^{\mu/\theta}\frac{V}{\lambda^3} \qquad (32)$$

Application of Eq (30) yields the usual results for ideal gases, such as $PV = \bar{N}kT$, however, the grand average of $\bar{N}$ enters the relations now rather than just N. From the definitions (27) and (28a) one may show that the fractional fluctuation in the number N is given by

$$\left(\frac{\overline{N_{gr}^2} - (\bar{N}_{gr})^2}{\bar{N}_{gr}^2}\right)^{1/2} = \frac{1}{\sqrt{\bar{N}_{gr}}} \qquad (33)$$

Thus for gases, the fluctuations in number are negligibly small, showing that the number of particles in the grand-canonical ensemble is sharply peaked around the average value. One therefore expects that the physical results deduced from the canonical and grand-canonical ensemble will be the same; the calculations, however, are frequently (especially in quantum problems) simpler in the grand canonical than in the canonical scheme.

Applications. Two of the newest and most important applications of statistical physics involve [illegible]

[3] Such as Eq. (15), which connects the thermodynamic free energy to the partition function, lies in the general validity of these connections. Specifically, when dealing with a gas of interacting molecules, where the energy is given by

$$E(x_1, \ldots, x_N; p_1, \ldots, p_N) = \sum_{i=1}^{N} \frac{p_{x_i}^2 + p_{y_i}^2 + p_{z_i}^2}{2m} + U(x_1, \ldots, x_N)$$

with the potential U a known function, the relation is still valid. In that case, one may obtain the partition function directly in terms of U, by performing the integrations over the momenta:

$$Z_N = \frac{1}{N!}\frac{1}{\lambda^{3N}}\int \cdots \int d^3x_1 \cdots d^3x_N \exp\left[-\frac{U(x_1, \ldots, x_N)}{kT}\right] \qquad (34)$$

Since from Z_N all thermodynamic relations follow, Eq. (34) is the appropriate starting point for all studies of nonideal gases. The canonical ensemble has been used in obtaining Z_N, which depends on V through the integration limits, on T through λ (see 31a), and on N through the number of integrations. Since Z_N gives the free energy immediately, one can get the equation of state (the relation between P, V, and T), by computing $P = -\partial\psi/\partial V$. The experimental results on the equations of state may be expressed in terms of the so-called virial expansion:

$$P = \frac{RT}{V}\left(1 + \frac{B(T)}{V} + \frac{C(T)}{V^2} + \cdots\right) \qquad (35)$$

Here $B(T)$ and $C(T)$ are the second and third virial coefficients; they are determined experimentally. It is clear that (35) is a development starting from the ideal gas equation. A first task for statistical mechanics would be to deduce the form (35), as well as an explicit expression for $B(T)$, etc.

$$\text{Let} \quad \exp\left[-\frac{U(x_1, \ldots, x_N)}{kT}\right] = W(1,2, \ldots, N) \qquad (36)$$

Assume now that the potential $U(x_1, \ldots, x_N)$ is such that it is possible to make an unambiguous distinction between interacting and noninteracting particles. Any potential which vanishes for a separation larger than a critical amount has this property, and any short-range potential approximates it. Call a separated configuration one in which none of the molecules of one group interacts with any of another. If two such groups are called α and β, one has, for such a configuration,

$$U = U(\alpha) + U(\beta) \qquad (36a)$$

hence, by (36),

$$W = W(\alpha)W(\beta) \qquad (36b)$$

Consider now a set of functions S defined by

$$W(1) = S(1) = 1$$
$$W(1,2) = S(1,2) + S(1)S(2)$$
$$W(1,2,3) = S(1,2,3) + S(1)S(2,3) + S(2)S(1,3) + S(3)S(1,2) + S(1)S(2)S(3)$$
$$W(1, \ldots, N) = S(1, \ldots, N) + S(1)S(2, \ldots, N) + S(1,2)S(3, \ldots, N) + \cdots + S(1) \cdots S(N) \qquad (36c)$$

From Eqs (36b) and (36c), one now proves the important property: S functions vanish for a sepa-

The canceled symbols in the numerator mean that one does not integrate over the coordinates and momenta of the molecule 1. The result given by Eq (19) is again true for an ideal gas. The identification of $\theta = kT$, however, is made in general. By differentiating Eq. (13) with respect to θ, one may establish the general connection (using $\theta = kT$)

$$\bar{E} = \psi - T\frac{\partial\psi}{T} = \epsilon \qquad \text{[by Eq. (18)]} \tag{20}$$

When this general relation, deduced from the canonical ensemble, is compared with the general thermodynamic relation (15d), one sees indeed that the parameter ψ and the free energy ψ' play identical roles. Again, for an ideal gas one can compute ψ, and it is the same as the ideal gas free energy. Hence one identifies ψ and ψ' in general. Equation (13) relates the thermodynamic free energy to an integral containing the mechanics of the microscopic problem. It is still interesting to observe that a quantity

$$\eta' = \frac{\int \rho \log \rho \, d\Gamma}{\int \rho \, d\Gamma} \tag{21}$$

plays in all instances the role of the entropy. In fact, one may show that $\psi = \epsilon - T\eta'$, in harmony with (15a). One can follow these thermodynamic analogies through in all detail to see that the canonical ensemble combined with the relevant definitions does indeed reproduce the formal aspects of thermodynamics The average energy $\bar{E}$ for a canonical ensemble has already been calculated. It is important to know just how often appreciable deviations from this average energy can be expected. For this, one needs to calculate the fractional fluctuations in energy given by

$$\frac{\overline{E^2} - (\bar{E})^2}{(\bar{E})^2}$$

The definition of $\overline{E^2}$ follows again from Eq (6):

$$\overline{E^2} = \frac{\int E^2 e^{-E/\theta} \, d\Gamma}{\int e^{-E/\theta} \, d\Gamma} \tag{22}$$

From Eq. (22) one may directly express $\overline{E^2}$ in terms of Z, the partition function For an ideal gas one can then obtain, using Eq. (17), an explicit expression for the fluctuations in energy:

$$\left(\frac{\overline{E^2} - (\bar{E})^2}{(\bar{E})^2}\right)^{1/2} = \sqrt{\frac{2}{3N}} \tag{23}$$

Since one deals with systems where the number of particles N is about 10^{22}, the chance of observing a sizable deviation from the average energy is extremely small. Using these same methods, defining in addition the specific heat $C_v = \partial\bar{E}/\partial T$, one may show generally that the fractional fluctuation in energy is given by

$$\left(\frac{\overline{E^2} - (\bar{E})^2}{(\bar{E})^2}\right)^{1/2} = \left(\frac{kT^2 C_v}{(\bar{E})^2}\right)^{1/2} \tag{24}$$

Even though the energy fluctuations are negligible for an ideal gas, this is not always so for other systems. For solids at low temperatures, the fluctuations may be appreciable. When phase transitions (of the first order) take place, C_v becomes infinite, indicating via Eq. (24) that the fluctuations become very large. It is clear that the main problem of the applications is reduced to the calculation of the partition function Z; all thermodynamic entities follow from Eqs. (16) and (13).

There is yet another ensemble which is extremely useful and which is particularly suitable for quantum mechanical applications. Much work in statistical mechanics is now based on the use of this so-called grand-canonical ensemble. In the grand ensemble one has a collection of systems; the number of particles in each system is no longer the same, but varies from system to system The density function $\rho(N,p,x)\ d\Gamma_N$ gives the probability that there will be in the ensemble a system having N particles, and that this system, in its $6N$-dimensional phase space Γ_N, will be in the region of phase space $d\Gamma_N$. The function ρ is given by

$$\rho(N,p,x) = \exp\left(\frac{\Omega + \mu N - E(p,x)}{\theta}\right) \tag{25}$$

Here θ, Ω, and μ are the parameters characterizing the ensemble, just as ψ and θ characterize the canonical ensemble. It is again true that these parameters are directly related to thermodynamic state functions. The detailed argument follows the identical pattern indicated in the discussion of the canonical ensemble The normalization condition is now

$$\sum_N \int d\Gamma_N \, \rho(N,p,x) = \sum_N \int d\Gamma_N \exp\left(\frac{\Omega + \mu N - E(p,x)}{\theta}\right) = 1 \tag{26}$$

The grand canonical average of any quantity Q is

$$\bar{Q}_{gr} = \sum_N \int d\Gamma_N \, \rho(N,p,x) Q_N(p,x) \tag{27}$$

and the grand partition function is

$$Z_{gr} = \sum_N e^{\mu N/\theta} \int d\Gamma_N \, e^{-(F/\theta)} \tag{28}$$

Again one customarily uses for a system consisting of N identical particles

$$Z_{gr} = \sum_N \frac{1}{N!} \frac{1}{h^{3N}} e^{\mu N/\theta} \int d\Gamma_N \, e^{-(E/\theta)} \tag{28a}$$

From Eq. (26) one sees immediately that

$$Z_{gr} = e^{-(\Omega/\theta)} \tag{29}$$

From Ω, sometimes called the grand potential, all other thermodynamic functions may be computed. The parameter μ is the chemical potential It is defined by $\mu = (\partial\xi/\partial N)_{P,T}$ where ξ is the Gibbs free

coordinates except those of molecule 1, one obtains an equation for f:

$$\frac{\partial f_1}{\partial t} + \frac{p_\alpha}{m_1}\frac{\partial f_1}{\partial x_\alpha} - \sum_{\alpha=1}^{3} \frac{\partial V_0}{\partial x_\alpha}\frac{\partial f_1}{\partial p_\alpha} = \iint d^3p_2\, d^3x_2 \sum_\alpha \frac{\partial U(x_1x_2)}{\partial x_\alpha}\frac{\partial f_2(p_1p_2,x_1x_2)}{\partial p_\alpha} \quad (44)$$

The equation for f_l *involves* f_{l+1}, etc.

The discussion of this hierarchy forms the basis of the study of the nonequilibrium phenomena in dense gases The basic difficulty is that the equations for f_l always involve a function f_{l+1}. To obtain an equation for a single f_l function, one needs to make a guess or assumption about the way in [illegible]

superposition approximation, which states that in some approximate sense

$$f_2(p_1x_1,p_2x_2) \cong f_1(p_1x_1)f_1(p_2x_2) \quad (45)$$

An appeal to probabilities of independent events would make Eq (45) appear reasonably plausible It should be stressed, however, that nowhere does one prove that the approximation of Eq (45) is indeed consistent with the system of Eq. (44) The use of Eq (45) in Eq (44) will lead to (nonlinear) integral equations In spite of the considerable amount of work done with these equations, [illegible]

[illegible] $U(1, \ldots, N) = \Sigma U(r_{ij})$, it is possible to express the equation of state in terms of the pair distribution function This function is defined by

$$n_2(x_1,x_2) = N(N-1)\frac{\int \cdots \int d(1)\, d(2)\, d(3) \cdots d(N)\, W(1, \ldots ,N)}{\int \cdots \int d(1) \cdots d(N)\, W(1, \ldots ,N)} \quad (46)$$

Here $W(1, \ldots, N)$ is defined by Eq. (36). The function n_2 gives the positional distribution of molecular pairs As such it is less detailed than f_2, which gives the distribution of pairs both in positions and momenta Apart from combinatorial factors, one has

$$n_2(x_1\,x_2) = \iint d^3p_1\, d^3p_2\, f_2(x_1p_1,x_2p_2) \quad (46a)$$

For central additive potentials, it may now be shown that n_2 is in fact a function of the distance r_{12} between the molecules, $n_2 \propto g(r)$, where $g(r)$ is defined as giving the number of molecules between r and $r + dr$

An analysis similar to the one given in the preceding discussion on nonideal gases now leads to an equation of state in terms of $g(r)$ alone

$$PV = NkT - \frac{2\pi N(N-1)}{3V}\int g(r)\frac{dU}{dr}r^3\,dr \quad (47)$$

The integral equation referred to for the f_2 functions (once the superposition approximation is made) may now be expressed as an approximate integral equation for $g(r)$. This equation, obtained by M. Born, H. S Green, and J. Yvon, leads to approximations for the virial coefficient, which are fairly good but *not excellent*. *It is important to recall that the pair distribution function* $g(r)$ may be obtained directly from experimental x-ray scattering data. In principle one could use experimental x-ray data to obtain via Eq. (47) data about the equation of state. For this to be a feasible procedure would demand accuracies much beyond the present limits, as well as extensive measurements over a range of temperatures and densities. Unfortunately, therefore, this is not a possible procedure at present (1959), although the formal relation (47) is generally valid In a current theory of *liquid helium*, one makes use of the relation between the pair distribution function and neutron-scattering data (*see* HELIUM, LIQUID).

Quantum statistical mechanics. The ensemble techniques can also be applied to systems which are described in terms of quantum mechanics. Consider an ensemble, where each system is described by a Hamiltonian operator H. Let $\psi^\alpha(x,t)$ be a wave function (x stands for $x_1, \ldots, x_N$, and α characterizes an ensemble member). The Schrödinger equation is

$$H\psi^\alpha(x,t) = -\frac{\hbar}{i}\frac{\partial\psi^\alpha(x,t)}{\partial t} \quad (48)$$

(*see* QUANTUM THEORY, NONRELATIVISTIC).

If $\varphi_n(x)$ is a complete orthogonal set,

$$\int\varphi_n{}^*\varphi_m = \delta_{nm} \quad (49)$$

the function ψ^α may be developed in terms of the set (φ)

$$\psi^\alpha(x,t) = \sum_n a_n{}^\alpha(t)\varphi_n(x) \quad (50)$$

Here $|a_n{}^\alpha(t)|^2$ is the probability that ensemble member α, at time t, is in state n. One has, of course, the fact that

$$\sum_n |a_n{}^\alpha(t)|^2 = 1 \quad (51)$$

Suppose the number of ensemble members is N_e. One defines now an ensemble average of a quantity G which is defined for each ensemble system by G^α as

$$\bar{\bar{G}} = \frac{1}{N_e}\Sigma(G^\alpha) \quad (52)$$

It is important to distinguish the ensemble average from the ordinary quantum mechanical average, which is defined for a single system:

$$\bar{G}^\alpha = \int(\psi^\alpha)^* G_{op}\psi^\alpha \quad (52a)$$

The notation, double bar for ensemble average, single bar for quantum mechanical average, stresses the difference. Of special importance in the calculation of averages is the density matrix The matrix elements of ρ, the density matrix, relative to the set of functions φ are defined by

rated configuration For instance $S(1,2, \ldots, l)$ is zero, unless the molecules 1, 2, . . . , l form a nonseparated configuration. This is an important property, for it allows the integration of the S functions in a simple fashion. Consider

$$\int \cdots \int d^3x_1 \cdots d^3x_l\, S(1, \ldots, l) \quad (37)$$

Imagine that molecule 1 is fixed somewhere in the middle of the vessel Since the S functions vanish for a separated configuration, the l molecules must *all be quite near to molecule 1. Roughly speaking,* if a is the range of the molecular forces, the integral (37) will get contributions only from a range of about la around molecule 1. If la is smaller than $V^{1/3}$, the first $l-1$ integrations will be independent of V. The last integrations over the coordinates of 1 will just contribute V, for molecule 1 can be placed anywhere in V. (Wall effects are ignored.)
Thus

$$\int \cdots \int dx_1 \cdots dz_l\, S(1, \ldots, l) = Vl!b_l(T) \quad (37a)$$

The $l!$ is a normalization factor and b_l is independent of V. The b_l are called cluster integrals. Using Eqs. (34), (36), (36c), and (37a), one may obtain the pressure in terms of the cluster integrals

$$P = \frac{RT}{V}\left(1 - \frac{Nb_2}{V} + 4(N^2b_2^2 - 2N^3b_3)\frac{1}{V^2} + \cdots\right)$$

One therefore has succeeded in deducing the experimental form of the virial development, while the cluster integrals are related to the experimental virial coefficients by

$$B(T) = -Nb_2 \quad (39a)$$
$$C(T) = N^2(4b_2^2 - 2b_3) \quad (39b)$$

Using (37a) for $l = 2$ gives a direct relation between B and the intermolecular force potential $U(r)$:

$$B(T) = 2\pi N \int_0^\infty r^2\, dr(1 - e^{-[U(r)/kT]}) \quad (39c)$$

This is a "perfect" statistical formula. In the case of helium, where both $U(r)$ and B are known, Eq (39c) may in fact be checked (At low temperatures, quantum effects, not included in Eq. (39c), begin to play an important role.)

One can obtain the complete equation of state using a similar procedure. The result comes out in an implicit form:

$$P = kT\Sigma b_l z^l \quad (40a)$$
$$\frac{N}{V} = \sum_{l=1}^{N} lb_l z^l \quad (40b)$$

These equations are due to J. E. Mayer. In principle, one can eliminate the auxiliary variable z between Eqs. (40a) and (40b) to obtain a relation between P, V, and T. To have a useful relation one must, of course, know the general character of the cluster integrals which, except for special molecular models, is not known. A question which has concerned physicists for some time is whether or not the system of equations (40) actually predicts a condensation phenomenon. It is well known that every real gas at a low enough temperature will condense when the volume is decreased. During the condensation process the pressure remains constant, the onset of condensation being marked by a discontinuity in the slope of the isotherm The problem is now whether this information can be obtained from Eqs. (40). An interesting clue was obtained by Mayer. He showed that for low enough temperatures, $b_l(T) \cong \text{constant} \times b_0{}^l$. Substituting in Eq. (40b), one obtains

$$\frac{N}{V} = \sum_{l=1}^{N} l(\text{constant})(b_0z)^N \quad (41)$$

Since the last power in the series is $N \cong 10^{24}$, one sees that the series given by Eq. (41) is radically different for $b_0z < 1$ and $b_0z > 1$. If $b_0z > 1$, $(b_0z)^N$ is a very large number and a change in N/V will cause a very slight change in z, hence a very slight change in P by Eq. (40a). Hence $b_0z = 1$ separates two ranges: if $b_0z < 1$, changes in N/V cause reasonable changes in P, if $b_0z > 1$, P becomes quite insensitive to changes in N/V. This qualitative idea has been refined in many ways, by rigorously studying the limit

$$\lim_{\substack{N\to\infty \\ V\to\infty}} (Z_N)^{1/N}$$

These studies have clarified some aspects of the situation, but even so the problem is still not completely settled.

Nonequilibrium theory of liquids. It was observed in the earlier discussion of Liouville's theorem that the Liouville equation usually does not help in the setting up of an appropriate ensemble. Yet the Liouville equation has become the starting point for an important development aimed at an understanding of the nonequilibrium states of dense gases and liquids This development starts from the Liouville equation

$$\frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N}\left(\frac{\partial \rho}{\partial x_i}\frac{dx_i}{dt} + \frac{\partial \rho}{\partial p_i}\frac{dp_i}{dt}\right) = 0 \quad (42)$$

The interpretation that $\rho(x_1, \ldots, p_{z_N}; t)\, d\Gamma$ is the probability of finding the molecules of the system in their prescribed momentum and position ranges may now be used Define a set of probability functions (or distribution functions) by

$$f_1(\mathbf{x}_1,\mathbf{p}_1,t) = \int \cdots \int d(1)\, d(2) \cdots dN\, \rho(x,p,t) \quad (43a)$$

$$f_l(\mathbf{x}_1, \ldots, \mathbf{p}_l; t) = \int \cdots \int d(1) \cdots d(l)\, d(l+1) \cdots d(N)\, \rho(x,p,t) \quad (43b)$$

One observes that f is (apart from constants) the Boltzmann distribution function. The canceled symbol $d(1)$ means that one is not to integrate over $x_1, y_1, z_1, p_{x_1}, p_{y_1}, p_{z_1}$. The higher distribution functions become important for dense systems. Assume that the potential energy of the system consists of an external potential V_0, and a potential U which is additive. Then by integrating Eq. (42) over all

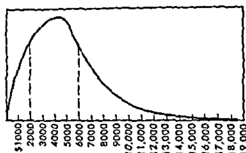


Fig 1. Distribution of incomes

Random variables are either continuous, which means they can take on any numerical value (the length of a room), or discrete, which means they can take on only a restricted set of values (number of windows in a room).

Distributions. In a univariate population, the population distribution is a curve (function of the random variable which characterizes the elements of the population) from which one can determine the proportion of the population which has elements in a certain range of the random variable For example, the curve of Fig 1 provides the distribution of annual incomes of family units in the United States in 1954 The total area under the curve is 1 The area under the curve between any two vertical lines gives the proportion of the families having annual incomes between the two values marked on the horizontal scale by the two vertical lines Thus, the fact that the area under the curve between $2000 and $6000 is 0 541 means that 54.1% of United States family units had incomes in that range in 1954

The distribution is also referred to as the distri-

... each point can also be plotted to give a curve which starts at zero and reaches unity as the variable becomes large, the resulting curve is sometimes called the cumulative distribution function, the probability distribution or simply the distribution The cumulative form of the curve of Fig 1 is shown in Fig 2, the height of the curve at any point on the horizontal scale equals the area to the left of that point under the curve of Fig 1 and is the proportion of the population having incomes less than the value at that point The distribution (in either frequency or cumulative form) gives com-plete ...

... are location parameters or measures of central tendency, a second class of ... parameters ...

... mean is the average over all the population of the values of the random variable It is often represented by the Greek letter μ. In mathematical terms, letting x be the random variable, $f(x)$ the frequency function for a given population, $F(x)$ its cumulative form, then

$$\mu = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-\infty}^{\infty} x\,dF(x)$$

The median, often designated by Med, M, or $X_{.50}$, is a number such that at most, one-half the values of the variable associated with the elements of the population fall above or below it

$$\int_{-\infty}^{M} dF(x) \geq \tfrac{1}{2} \leq \int_{M}^{\infty} dF(x)$$

The mode is the most frequent value of the random variable; if the frequency function has a unique maximum value, the mode is the value of the random variable at which the frequency function reaches its maximum. Location parameters are numbers near the center of the range over which the random variable of the population varies; different ones arise from different definitions of center, generally the mean is used unless special circumstances make some other location parameter more appropriate.

The extent to which a population is scattered on either side of its center is roughly indicated by measures of dispersion such as the standard deviation, the mean deviation, the interquartile range, the range, and sometimes others The standard deviation is the square root of the mean square of the deviations from the mean, it is usually denoted by the Greek letter σ, σ^2 is called the variance.

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \int_{-\infty}^{\infty} (x-\mu)^2\,dF(x)$$

The mean deviation is the average over the population of the deviations from the mean, all taken to be positive.

$$\text{Mean deviation} = \int_{-\infty}^{\infty} |x-\mu| f(x)\,dx = \int_{-\infty}^{\infty} |x-\mu|\,dF(x)$$

The interquartile range (often denoted by Q) is the difference $X_{.75} - X_{.25}$, where $X_{.75}$ is the value of the random variable such that one-quarter of the population has values larger than $X_{.75}$, and $X_{.25}$

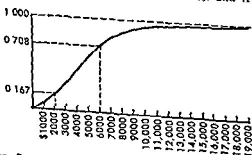


Fig 2. Cumulative distribution of incomes

$$\rho_{mn} = \frac{1}{N_e} \sum_{\alpha} (a_n^{\alpha})^* a_m^{\alpha} = \overline{\overline{a_n^* a_m}} \qquad (53)$$

One sees that the sum of the diagonal elements of ρ, called the trace of ρ, is

$$\text{Tr}(\rho) = \sum_n \rho_{mn} = \frac{1}{N_e} \sum_{\alpha} \sum_n |a_n^{\alpha}|^2 = 1 \qquad (53a)$$

The density matrix ρ is the counterpart of the classical density function. From the Schrödinger equation (50) one deduces immediately the analog of the Liouville theorem

$$i\hbar \frac{d\rho_{mn}}{dt} = [H,\rho]_{mn} \qquad (54)$$

Here $[a,b] \equiv ab - ba$ is the commutator. Finally one obtains the basic relation, giving the ensemble average of any operator as

$$\overline{\overline{G}} = \text{Tr}(\rho G) \qquad (55)$$

The fact that observable entities, such as ensemble averages, come out in the form of a trace, implies that the calculated results are independent of the set of functions φ. This is true because a change in this basic set will induce a similarity transformation on both ρ and G, and the trace operation is invariant under such a transformation.

If one has an ensemble in which all ensemble members are in the identical state (ψ^{α} is independent of α), one has a pure case. If the ψ^{α} do depend on α, one has a statistical mixture. A necessary and sufficient condition for a pure case is that $\rho^2 = \rho$.

At equilibrium, one has the counterpart of the canonical ensemble. Then the density operator, see Eq. (12), is written as

$$\rho = e^{(\psi - H)/kT} \qquad (56a)$$

The free energy is again directly related to the partition function

$$Z = e^{-(\psi/kT)} = \text{Tr}(e^{-(H/kT)}) \qquad (56b)$$

In Eq. (56), where Q is an operator e^Q defined by the series expansion

$$e^Q = \sum_n \frac{Q^n}{n!} \qquad (57)$$

The thermodynamic relations follow as before.

The construction of an ensemble in quantum mechanics must be performed in harmony with the knowledge of the system. If the system is known to be in either one of two quantum states, the ensemble members must be evenly distributed over these states, with, in addition, random phases. It is important to include the requirement of random phases; if this were not done a uniform stationary ensemble would not even remain stationary.

[M DR.]

Bibliography: T. L. Hill, *Statistical Mechanics*, 1956; J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids*, 1954; A. Khinchin, *Mathematical Foundations of Statistical Mechanics*, 1949; L. D. Landau and E. Lifshitz, *Statistical Physics*, 1958; J. E. Mayer and M. G. Mayer, *Statistical Mechanics*, 1940; R. C. Tolman, *The Principles of Statistical Mechanics*, 1938.

Statistics

The field of knowledge concerned with collecting, analyzing, and presenting data. Not only workers in the physical, biological, and social sciences, but also engineers, business managers, government officials, market analysts, and many others regularly use statistical methods to greater or lesser degree in their work. The methods range from simple counting to complex mathematical systems designed to extract the maximum amount of information from very expensive data.

In an important sense statistics may be regarded as a field of application of probability theory. The common problem faced by a physicist reading a meter, an engineer testing a material, an agronomist measuring the yield of a hybrid corn, a chemist determining the concentration of ascorbic acid, and an interviewer studying public opinion is the problem of random variation which prevents repetition of exactly the same result when a measurement is repeated. Statistical methods are employed to assess the magnitude of random variation, to minimize it, to balance it out, to remove it by calculation procedures, and to analyze it by suitably arranged patterns of observation. The theory of probability is concerned with the properties of random variables and hence furnishes the basis for developing techniques for controlling them. See PROBABILITY

Viewing statistics from another direction, it is the science of deriving information about populations by observing only samples of those populations. A population is any well-specified collection of elements. Thus, one may refer to the population of adults in the continental United States viewing television screens at 8:14 P.M. on August 6, 1960; the population of automobiles less than two years old registered in Los Angeles County on a certain date; the population of vineyards in France; the hypothetical population of outcomes of tossing a given coin endlessly. Populations may be finite or infinite. An element of a univariate population is characterized by the value of a random variable which measures some single attribute of interest in the population. Thus, one may be interested in whether or not individuals of the television audience were or were not viewing program A, with each individual one may associate a random variable, let it be X, which takes on the value of 1 if the individual is watching A and 0 if he is not. If one were interested in a second characteristic of the elements of the television audience (such as age), he would be said to be dealing with a bivariate population; a third characteristic (such as economic status) would make it a trivariate or, less specifically, a multivariate population.

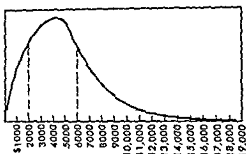


Fig 1. Distribution of incomes.

Random variables are either continuous, which means they can take on any numerical value (the length of a room), or discrete, which means they can take on only a restricted set of values (number of windows in a room)

Distributions. In a univariate population, the population distribution is a curve (function of the random variable which characterizes the *elements* of the population) from which one can determine the proportion of the population which has elements in a certain range of the random variable. For example the curve of Fig 1 provides the distribution of annual incomes of family units in the United States in 1954. The total area under the curve is 1 The area under the curve between any two vertical lines gives the proportion of the families having annual incomes between the two values marked on the horizontal scale by the two vertical lines Thus, the fact that the area under the curve between $2000 and $6000 is 0 541 means that 54.1% of United States family units had incomes in that range in 1954

The distribution is also referred to as the distribution function, the density function, the frequency function, or the probability density.

The total area under the distribution curve to the left of each point can also be plotted to give a curve which starts at zero and reaches unity as the variable becomes large, the resulting curve is sometimes called the cumulative distribution function the probability distribution or simply the distribution The cumulative form of the curve of Fig 1 is shown in Fig 2 the height of the curve at any point on the horizontal scale equals the area to the left of that point under the curve of Fig 1 and is the proportion of the population having incomes less than the value at that point The distribution (in either frequency or cumulative form) gives complete information about the way the characterizing variable is spread through the population

Population parameters. Populations (or population distributions) are often specified incompletely by certain population parameters Some of these parameters are location parameters or measures of central tendency; a second class of important parameters consists of measures of dispersion or scale parameters

The most widely used location parameters are the mean, the median, and the mode The mean is the average over all the population of the values of the random variable It is often represented by the Greek letter μ In mathematical terms, letting x be the random variable, $f(x)$ the frequency function for a given population, $F(x)$ its cumulative form, then

$$\mu = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-\infty}^{\infty} x\,dF(x)$$

The median, often designated by Med, M, or $X_{.50}$, is a number such that at most, one half the values of the variable associated with the elements of the population fall above or below it

$$\int_{-\infty}^{M} dF(x) \geq \tfrac{1}{2} \leq \int_{M}^{\infty} dF(x)$$

The mode is the most frequent value of the random variable; if the frequency function has a unique maximum value, the mode is the value of the random variable at which the frequency function reaches its maximum. Location parameters are numbers near the center of the range over which the random variable of the population varies, different ones arise from different definitions of center, generally the mean is used unless special circumstances make some other location parameter more appropriate.

The extent to which a population is scattered on either side of its center is roughly indicated by measures of dispersion such as the standard deviation, the mean deviation, the interquartile range, the range, and sometimes others The standard deviation is the square root of the mean square of the deviations from the mean; it is usually denoted by the Greek letter σ; σ^2 is called the variance.

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \int_{-\infty}^{\infty} (x-\mu)^2\,dF(x)$$

The mean deviation is the average over the population of the deviations from the mean, all taken to be positive.

$$\text{Mean deviation} = \int_{-\infty}^{\infty} |x-\mu| f(x)\,dx = \int_{-\infty}^{\infty} |x-\mu|\,dF(x)$$

The interquartile range (often denoted by Q) is the difference $X_{.75} - X_{.25}$, where $X_{.75}$ is the value of the random variable such that one-quarter of the population has values larger than $X_{.75}$, and $X_{.25}$

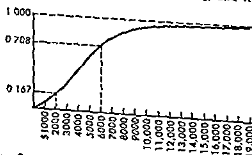


Fig 2 Cumulative distribution of incomes

is the number such that one-quarter of the population has values smaller than X_{25}. The three numbers, X_{25}, X_{50}, X_{75}, are called quartiles; these divide the population into quarters. The range is the difference between the largest and the smallest of the population elements.

Samples. If one examines every element of a population and records the value of the random variable for each, then he has complete information about the distribution of the random variable in the population, and there is no statistical problem.

It is usually impossible or uneconomical to make a complete enumeration (or census) of a population and one must therefore be content to examine only a part or sample of the population. On the basis of the sample, one draws conclusions about the entire population; the conclusions thus drawn are not certain in the sense that they would likely have been somewhat different if a different sample of the population had been examined. The problem of drawing valid conclusions from samples and of specifying their range of uncertainty is known as the problem of statistical inference.

Statisticians distinguish two kinds of samples. A survey sample is one chosen from a population, all the elements of which actually exist. An experiment is a sample chosen from a hypothetical population. Thus if one were interested in the total number of pigs being fattened in a given season by the farmers of the state of Iowa, he would survey-sample

periment by giving a group of pigs the hormone for a period of time; using the sample of observations thus obtained from the experiment he would draw inferences about the hypothetical population of observations that would have resulted had all pigs been given the hormone.

Random sampling. In planning a sample survey, the manner in which the data to be gathered will fulfill the purpose should be clearly stated. The population to be sampled must be explicitly defined. The method of sampling should be efficient and lead to a straightforward analysis. The question of what elements should be included in the population depends on the purpose of the survey. Thus the population of vineyards in France might include as a vineyard a dozen vines in the back yard of a man living in the heart of Paris if the purpose were to estimate total grape production of France; but the population might be defined to exclude such a small vineyard if the purpose were to estimate the size of the harvest to be available to commercial wineries.

The type of sampling of interest in statistics is probability sampling, because it eliminates subjective aspects from the selection of the sample. In probability sampling, all possible distinct samples are known, the selection of the sample is done randomly according to a preassigned probability, and the method of analysis is predetermined and unambiguous. Only from such samples can inferences about populations be made with measurable precision.

Simple random sampling is a method of selecting a sample of n elements out of a population of N elements so that all such samples have an equal probability of being drawn. This may be done by selecting a first element at random from the population, then a second element at random from the remaining population, and so on until the n elements are selected. Because an element cannot appear more than once in the sample, this is a form of sampling without replacement. The sampling ratio or sampling fraction is n/N.

Often the purpose of drawing a sample is to estimate the mean or average of a characteristic of the population. If y_i is the value of the characteristic of the ith unit, then the population mean is

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{N}(y_1 + y_2 + y_3 + \cdots + y_N)$$

The sample mean is the average of the n units in the sample:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \qquad + x_n)$$

where the n x's are the n y's selected from the population as the sample. The population total is $N\mu$ and is estimated by $N\bar{x}$. The population variance is defined as

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (y_i - \mu)^2$$

and the sample variance is usually defined as follows, although some authors use n instead $n - 1$:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

The quantities $\bar{x}$ and s^2, the sample mean, and the sample variance are called sample statistics, they are the first and second sample moments and are also estimators of the corresponding population parameters, μ and σ^2.

The estimators are themselves random variables. If one repeatedly drew samples of size n from the population and computed $\bar{x}$ from each sample he would obtain a population of $\bar{x}$'s with its own distribution which would differ from the distribution of x. It can be shown that the $\bar{x}$ population has exactly the same mean μ as the x population. Further, the variance of the $\bar{x}$ population is

$$\frac{(N - n)}{(N - 1)}\frac{\sigma^2}{n}$$

where σ^2 is the variance of the x population. The first fraction is ordinarily nearly unity so that the variance of $\bar{x}$ is approximately the fraction $1/n$ of the original population variance. As n, the sample size, becomes large, the $\bar{x}$ population becomes more

concentrated about μ and the reliability of a particular value of $\bar{x}$ as an estimate of the population mean increases

The observations of a sample, besides providing estimates of population parameters, can also be used to obtain an estimate of the population's frequency function. This estimate is determined by dividing the range of the sample observations into several intervals of equal length L and counting the number of observations occurring in each interval; these numbers are then divided by nL to determine fractions giving the relative density of the sample occurring in each interval; then on a sheet of graph paper one lays out the intervals on a horizontal axis and plots horizontal lines above each interval at a height equal to the fraction corresponding to the interval; finally the successive plotted horizontal lines are connected by vertical lines to form a broken line curve known as a histogram (Fig 3). The area under the curve is unity, and the area between any two points gives the fraction of the sample observations lying between those two points If one takes larger and larger samples the chosen intervals can be made smaller and the broken line curve will come closer and closer to the underlying population frequency function Often one does not trouble to divide the interval frequencies by nL to normalize the area of the histogram, but merely plots the frequencies themselves; the resulting broken line curve is still referred to as a histogram

Sampling techniques When a population can be regarded as being made up of several nonoverlapping subpopulations, one may draw a sample from it by drawing a simple random sample from each subpopulation (or stratum), this procedure is called stratified random sampling The method is emplo[illegible]
forma[illegible]
is adn[illegible]
ulation, when there are natural strata, or when a gain in precision would be realized because each stratum is more homogeneous than the whole population

Let n be the number of units sampled in the entire population of N units, and let n_h be the number of units sampled in stratum h containing N_h units Therefore, the sum of all the n_h is n, and of all the N_h is N Let σ_h be the true standard deviation within stratum h When $n_h/n = N_h/N$, the n_h are said to be proportionately allocated When $n_h/n = N_h\sigma_h/\Sigma N_h\sigma_h$, the n_h are said to be optimally allocated because the variance of the estimated mean is then smallest for fixed total size of sample. Ordinarily, if the σ_h are well estimated, optimum allocation gives greater precision than proportional allocation, and proportional allocation gives greater precision than simple random sampling.

Systematic sampling may be regarded as a sampling of units at regular intervals in the population, for example, every tenth unit. Usually this is done with a random start; that is, the first unit in the sample is selected at random from a small group at the beginning A systematic sample is usually relatively easy and fast to execute, and, if the analogy to stratified sampling is valid, a systematic sample will be more accurate. However, because of the essentially nonrandom nature of a systematic sample, it is difficult to estimate the sample variance unless certain assumptions are made about randomness in the order of the population If the population has any periodicity, then the estimates made from a systematic sample can be quite poor.

When the elements of a population are in mutually exclusive groups called the primary units of the population, then a sample of these primary units might be made, and then, from those selected, a sample of the individual elements would be made This procedure is called two-stage sampling or subsampling Multistage sampling can involve more than two stages of sampling. Although more complicated and difficult to apply and to analyze, multistage sampling offers considerable flexibility for balancing between statistical precision and the cost of sampling.

When too little is known about a population to plan a sample, then two samples are made The manner in which the second sample is taken is determined from the results of the first sample In Stein's method of two-stage sampling, the size of the additional sample is decided by using the estimate of the variance from the first sample In double sampling or two-phase sampling, the first sample is used to gather information on a variate associated with the variate of interest. This information is used to establish strata for drawing a stratified random sample involving the variate of interest

Ratio and regression estimates are often used when observations are made on one or more variates related to the variate of primary interest.

Sequential sampling and sequential analysis refer to a method of sampling in which the size of the sample is not specified in advance Observations are drawn one by one until a specified degree of confidence in the information to be obtained has been achieved

Sampling distributions. Many [illegible]

[illegible] symmetrical distribution centered at its mean μ. The distribution is illustrated in Fig. 4 for three different values of σ, the



Fig 3 Histogram

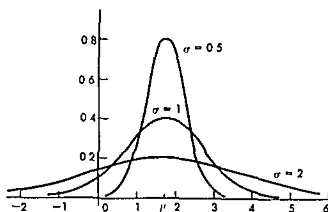


Fig. 4. Normal distribution

standard deviation The mathematical equation for these curves is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/\sigma^2}$$

The area under the curves is given below for the indicated limits

Area	*Limits*
0 50	$\mu - 0.675\sigma$ to $\mu + 0.675\sigma$
0 683	$\mu - \sigma$ to $\mu + \sigma$
0.90	$\mu - 1.645\sigma$ to $\mu + 1.645\sigma$
0 95	$\mu - 1.960\sigma$ to $\mu + 1.960\sigma$
0.99	$\mu - 2.326\sigma$ to $\mu + 2.326\sigma$

Referring to the above limits, 0.675σ is called the probable error inasmuch as the odds are even that a randomly drawn observation will lie within 0.675σ of the mean.

If samples of size n are drawn from a normal population and the sample mean $\bar{x}$ is computed for each, the $\bar{x}$'s will have a normal distribution with the same mean and with variance σ^2/n Thus, if samples of size 4 were drawn from a population distributed by the curve marked $\sigma = 1$ in Fig 4, then the means of those samples would be distributed by the curve marked $\sigma = 0.5$ (that is, $1/\sqrt{4}$) Further, the probability would be 0 95 that a particular $\bar{x}$ so drawn would lie within (1 960) (0 5) = 0.98 of the population mean μ.

The central limit theorem of the theory of probability states that under very general conditions the sample mean $\bar{x}$ is approximately *normally distributed* whatever may be the distribution function for the underlying population This powerful theorem enables one to make probability statements such as the one at the end of the preceding paragraph in ignorance of the actual population distribution.

Besides the distribution of the mean, two other sampling distributions derived from the normal *distribution* have *wide* application. One is the chi-square distribution which provides the distribution of the sample variance:

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1)$$

where $x_1, x_2, \ldots, x_n$ are the observations of a random sample of size n drawn from a normal population. The quantity

$$\chi^2 = (n-1)s^2/\sigma^2$$

has a distribution illustrated in Fig 5 for three values of n. The mathematical form for the distribution curve is

$$\frac{1}{\left(\frac{n-3}{2}\right)!}\left(\frac{n-1}{2\sigma^2}\right)^{\frac{n-1}{2}} (\chi^2)^{\frac{n-3}{2}} e^{-(n-1)\chi^2/2\sigma^2}$$

and it is often referred to as the chi-square distribution with $n - 1$ degrees of freedom.

The most useful random variable for interval *estimation of a population mean is*

$$t = \frac{\sqrt{n}(\bar{x} - \mu)}{s}$$

which has a symmetrical distribution very similar in appearance to the curves plotted in Fig 4 The mathematical form for the distribution is

$$\frac{[(n-2)/2]!}{\sqrt{(n-1)\pi}[(n-3)/2]!} \bigg/ \left(1 + \frac{t^2}{n-1}\right)^{n/2}$$

This is referred to as the t distribution or the Student distribution with $n - 1$ degrees of freedom The following table gives limits which include 95% of the area under the curve for a few values of n

n	*Limits*
2	−12.71 to +12.71
3	−4 30 to +4.30
4	−3.18 to +3 18
10	−2 20 to +2 20
25	−2.06 to +2.06
120	−1 98 to +1 98

For very large sample sizes, the limits become −1.96 to +1 96 as for the normal distribution. An illustration of the use of these limits is given below.

Estimation. In making an estimate of the value of a parameter of a population from a sample, a function (called the estimator) of the observations is used For example, for estimating the mean of a normal population the mean of the sample ob-

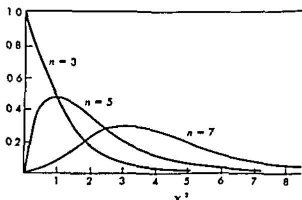


Fig 5. Chi-square distributions

servations is usually taken as the estimator. Another estimator is the average of the two most extreme observations. In fact, there is an infinitude of *estimators*. The problem of estimation is to find a "good" estimator.

A good estimator may be regarded as one which results in a distribution of estimates concentrated near the true value of the parameter and which can be applied without excessive effort. There is no single way of deciding how good an estimator is, but there are several criteria by which an estimator may be judged.

An unbiased estimator is one which results in a *distribution* of estimates which has a mean exactly equal to the value of the parameter being estimated. Otherwise the estimator is called biased. The bias is the mean of the distribution of the estimator minus the value of the parameter it estimates.

An estimator is said to be consistent if the probability that an estimate will differ from the value of the parameter by more than any fixed amount can be made arbitrarily small by increasing the number of observations.

The variance of an estimator is the mean squared deviation of the estimates from the value of the parameter. The estimator with the smallest variance is called most efficient. The relative efficiency of two estimators is the ratio of the variances. When the numerator of this ratio is the variance of a most efficient estimator, this ratio is simply called the efficiency of the other estimator.

An estimator is said to be sufficient if it contains all the information in the sample regarding the parameter. This is so when the conditional distribution of the sample for a given estimate is independent of the parameter.

There are several methods of constructing estimators of parameters. The method of moments is applied by assuming that the first few *sample* moments are equivalent to the moments of some distribution, and then solving for the parameters of that distribution. The methods known as least squares, minimum variance, and minimum chi-square all have as their basis the estimation of the values of the parameters which *minimize* some linear function of the squares of deviations of the observations from the values of the parameters. In applying Bayes' method, the distribution of possible values of the parameter before the sample is taken, called the a priori distribution, is used in conjunction with the observations in the sample to yield an estimator. The method of maximum likelihood uses as the estimate that value of the parameter for which the probability of the sample is highest.

A confidence interval is an interval constructed in such a way that the true parameter value is within this interval with a predeterm [illegible]

ity in repea [illegible]

the confider [illegible]

sample mea [illegible]

with variar [illegible]

the true but unknown mean, μ, with probability 0.95. Thus

$$\text{Probability that } (-1.96\sigma < \bar{x} - \mu < 1.96\sigma) = 0.95$$

On solving the two inequalities for μ, this expression may be written

$$\text{Probability that } (\bar{x} - 1.96\sigma < \mu < \bar{x} + 1.96\sigma) = 0.95$$

When a particular value of $\bar{x}$ is computed from the observations, then $\bar{x} \pm 1.96\sigma$ is a pair of numbers which define a particular interval called the confidence interval. Because μ is a fixed number, it is within this particular interval with probability zero or one. However, μ is unknown, so that the confidence to be associated with the interval is stated in terms of the proportion of all intervals constructed in this manner which would include μ, rather than this particular interval. A probability of this sort, valid for a population of outcomes, when used in reference to a particular outcome is called a fiducial probability.

Ordinarily σ is unknown and the estimate s derived from the sample must be used to form a confidence interval; in this case, the t distribution rather than the normal distribution must be used. As an example, suppose a chemist has made four determinations of *the atomic weight μ of hydrogen* as follows: 1.0066, 1.0090, 1.0084, and 1.0086. The average of these values is

$$\begin{aligned}\bar{x} &= \tfrac{1}{4}(1.0066 + 1.0090 + 1.0084 + 1.0086)\\ &= 1.0082\end{aligned}$$

and the sample estimate of the variance of his technique is

$$\begin{aligned}s^2 &= \tfrac{1}{3}\{(1.0066 - 1.0082)^2 + (1.0090 - 1.0082)^2\\ &\quad + (1.0084 - 1.0082)^2 + (1.0086 - 1.0082)^2\}\\ &= 0.00000123\end{aligned}$$

with the estimate of σ being therefore

$$s = \sqrt{0.00000123} = 0.0011$$

It follows from the definition of t in the preceding section that in this case

$$t = 2(1.0082 - \mu)/0.0011$$

and using the short t table given in the preceding section one finds

$$\text{Probability}\left(-3.18 < \frac{2(1.0082 - \mu)}{0.0011} < 3.18\right) = 0.95$$

Solving these inequalities for μ one finds

$$\text{Probability } (1.00645 < \mu < 1.00995) = 0.95$$

and the chemist can assert with 95% *confidence* that the atomic weight of hydrogen lies between 1.00645 and 1.00995. For greater precision, *more* observations are required.

In a similar fashion one may obtain a confidence interval for σ by using s^2 in connection with the chi-square distribution.

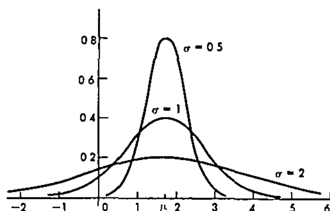


Fig. 4. Normal distribution.

standard deviation. The mathematical equation for these curves is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/\sigma^2}$$

The area under the curves is given below for the indicated limits.

Area	*Limits*
0.50	$\mu - 0.675\sigma$ to $\mu + 0.675\sigma$
0.683	$\mu - \sigma$ to $\mu + \sigma$
0.90	$\mu - 1.645\sigma$ to $\mu + 1.645\sigma$
0.95	$\mu - 1.960\sigma$ to $\mu + 1.960\sigma$
0.99	$\mu - 2.326\sigma$ to $\mu + 2.326\sigma$

Referring to the above limits, 0.675σ is called the probable error inasmuch as the odds are even that a randomly drawn observation will lie within 0.675σ of the mean.

If samples of size n are drawn from a normal population and the sample mean $\bar{x}$ is computed for each, the $\bar{x}$'s will have a normal distribution with the same mean and with variance σ^2/n. Thus, if samples of size 4 were drawn from a population distributed by the curve marked $\sigma = 1$ in Fig. 4 then the means of those samples would be distributed by the curve marked $\sigma = 0.5$ (that is, $1/\sqrt{4}$). Further, the probability would be 0.95 that a particular $\bar{x}$ so drawn would lie within (1.960) (0.5) = 0.98 of the population mean μ.

The central limit theorem of the theory of probability states that under very general conditions the sample mean $\bar{x}$ is approximately normally distributed whatever may be the distribution function for the underlying population. This powerful theorem enables one to make probability statements such as the one at the end of the preceding paragraph in ignorance of the actual population distribution.

Besides the distribution of the mean, two other sampling distributions derived from the normal distribution have wide application. One is the chi-square distribution which provides the distribution of the sample variance:

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1)$$

where $x_1, x_2, \ldots, x_n$ are the observations of a random sample of size n drawn from a normal population. The quantity

$$\chi^2 = (n-1)s^2/\sigma^2$$

has a distribution illustrated in Fig. 5 for three values of n. The mathematical form for the distribution curve is

$$\frac{1}{\left(\frac{n-3}{2}\right)!}\left(\frac{n-1}{2\sigma^2}\right)^{\frac{n-1}{2}} (\chi^2)^{\frac{n-3}{2}} e^{-(n-1)\chi^2/2\sigma^2}$$

and it is often referred to as the chi-square distribution with $n - 1$ degrees of freedom.

The most useful random variable for interval estimation of a population mean is

$$t = \frac{\sqrt{n}(\bar{x} - \mu)}{s}$$

which has a symmetrical distribution very similar in appearance to the curves plotted in Fig. 4. The mathematical form for the distribution is

$$\frac{[(n-2)/2]!}{\sqrt{(n-1)\pi}[(n-3)/2]!} \Big/ \left(1 + \frac{t^2}{n-1}\right)^{n/2}$$

This is referred to as the t distribution or the Student distribution with $n - 1$ degrees of freedom. The following table gives limits which include 95% of the area under the curve for a few values of n.

n	*Limits*
2	−12.71 to +12.71
3	−4.30 to +4.30
4	−3.18 to +3.18
10	−2.20 to +2.20
25	−2.06 to +2.06
120	−1.98 to +1.98

For very large sample sizes, the limits become −1.96 to +1.96 as for the normal distribution. An illustration of the use of these limits is given below.

Estimation. In making an estimate of the value of a parameter of a population from a sample, a function (called the estimator) of the observations is used. For example, for estimating the mean of a normal population the mean of the sample ob

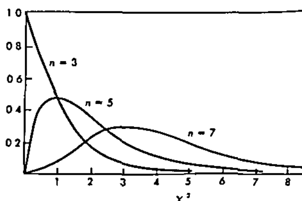


Fig. 5. Chi square distributions.

location	yes	no	undecided
east	n_{11}	n_{12}	n_{13}
south	n_{21}	n_{22}	n_{23}
midwest	n_{31}	n_{32}	n_{33}
west	n_{41}	n_{42}	n_{43}

Fig 7. Contingency table

the jth column of the table The sum of all the n_{ij} is, of course, n, the sample size. The row sums may be denoted by r_1, r_2, r_3, r_4, and the column sums by c_1, c_2, c_3; the sum of each of the sets is n On applying the likelihood criterion to the test of the hypothesis, one finds that it rests on the expression

$$2(n \log n + \sum_{ij} n_{ij} \log n_{ij} - \sum_i r_i \log r_i - \sum_j c_j \log c_j)$$

which is approximately distributed according to the chi-square distribution with 6 degrees of freedom To test at the 0.05 level for the type I error, [illegible]

if the expression turned out to be less than 12 6, one would accept the hypothesis of independence, otherwise he would reject it

Had there been R rows and C columns instead of four and three one would have used the chi-square distribution with $(R-1)(C-1)$ degrees of freedom For a trivariate population there would have been a three-way contingency table and four hypotheses of independence that could have been tested three of them assert that a given criterion of classification is independent of the other two, and the fourth asserts that all three are mutually independent

Design of experiments An experiment is performed to obtain information about the relations between several variables For example, one may study the effect of storage temperature and duration of storage on the flavor of a frozen food. Three variables (flavor temperature, and duration) are involved one (flavor) is called the subject of the experiment the other two are called factors which influence the subject. Sometimes the factors have intrinsic value in themselves, sometimes they are merely nuisance variables which must be taken into account because it is impossible to [illegible]

[illegible] experimental observations, the actual construction of the designs requires quite advanced mathematics based on group theory, finite geometries, and combinatorial analysis [illegible]

[illegible] designs are most important to the experimenter when observations are expensive to make and when more than one factor is involved in the experiment. In the past it was believed that the best experimental procedure was to vary factors one at a time. Thus, in the frozen food experiment one might have held storage temperature constant and studied the effect of duration of storage only on flavor. Having determined that relationship, one would then hold duration constant and study the temperature effect. This procedure is not only wasteful of time and resources but may very well lead to erroneous conclusions because in all likelihood there is interaction between the two factors; that is, duration effect probably changes in a not obvious way when temperature is changed. Even if one believes that he can extrapolate accurately the duration effect to other temperatures, there is much to be said for checking the extrapolation procedure in the experiment because it can probably be done at no additional cost if the experiment is well designed.

As an illustration of the use of a design, consider a very simple and useful one called the randomized block design Suppose a manufacturer contemplates purchasing a machine which can be obtained from one of three sources, X, Y, or Z He obtains one of each on trial in order to compare their performance using several operators from his own plant Perhaps five operators, A, B, C, D, and E, are to be used in the proposed experiment, which is a two-factor experiment with performance being the subject and the two factors being men and machines The operators are not really a factor of interest in evaluating the machines but are, of course, necessary to the experiment.

It would be a mistake to use perhaps 15 operators, putting 5 on each machine, because operators differ in ability and a machine that turned out well might have been fortunate in having a good set of operators assigned to it It is necessary that the machines be compared in as nearly equivalent circumstances as possible; in this case, that is done by having each operator operate every machine The machines may then be compared in blocks (operators) which are individually homogeneous although they may be quite different among themselves

The data are obtained by measuring production of each operator on each machine for a specified period of time and filling in the table of Fig 8 It is essential that the order in which a man works on the three machines be randomized (by tossing

operator \ machine	X	Y	Z
A			
B			
C			
D			
E			

Fig 8 Data form for randomized block [illegible]

A confidence region is the generalization of a confidence interval and refers to the simultaneous estimation of several population parameters. The confidence level is the proportion of the time that the region actually includes the true values of the parameters.

In general, the most desirable confidence interval or region is the smallest one which can be constructed for the selected confidence level.

Tests of hypotheses. Besides estimation of parameters, another major area of statistical inference is the testing of hypotheses. A hypothesis is merely an assertion that a population has a specific property. The test consists of drawing a sample from the population and determining whether or not it is consistent with the assertion. Very often the hypothesis is a statement about the mean of a population; that it has a given value, that it is the same as that of another population, that it exceeds that of another population by at least ten units, and the like Thus, one may be comparing a new blend of gasoline with a current blend, a new drug with a standard one, a new manufacturing process with an existing one

For a very simple illustration of the basic ideas involved, suppose a population is known to have either of two distributions which differ mainly in location. Let the random variable be x and let the two functions of x which determine the two distributions be represented by the symbols $f(x)$ and $g(x)$ as indicated in Fig. 6; that is, $f(x)$ represents the height of the left curve at x, and $g(x)$ represents the height of the right curve at x Assume that the population has the distribution $f(x)$; there is in this case only a single alternative, $g(x)$, so that rejection of the hypothesis implies acceptance of the alternative. Suppose that for testing the hypothesis one can afford only a single observation, that is, a sample of size one

It is obvious in this instance that one should choose some number b in advance, then accept the hypothesis if the observation falls to the left of b and reject it if the observation falls to the right Values of x to the left of b are said to constitute the acceptance region of the test, values to the right of b constitute the critical region. The area under $f(x)$ to the left of b, which is denoted mathematically by

$$\int_{-\infty}^{b} f(x)\,dx$$

is the probability that the hypothesis will be accepted when it is true The area to the right of b under $f(x)$ is the probability that the hypothesis will be rejected when it is true; this probability is

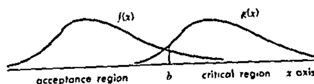


Fig 6 Simple test of a hypothesis.

called the type I error of the test. The area to the left of b under $g(x)$ is the probability that the hypothesis will be accepted when it is false and is called the type II error of the test. The area to the right of b under $g(x)$ is called the power of test and is, of course, one minus the type II error.

By moving the point b to the right one can make the type I error as small as he likes, but in doing so the type II error is increased This dilemma is characteristic of the construction of tests of hypotheses. Ordinarily one arbitrarily chooses the type I error to be some small number such as 0 05 or 0 01 and then chooses the critical region so as to minimize the type II error

The fraction $f(x)/g(x)$ is called the likelihood ratio or simply the likelihood function of x. From Fig. 6 it is evident that the likelihood is relatively large when the hypothesis is true and small when it is false. For a sample of size n with observations $x_1, x_2, \ldots, x_n$, the sample likelihood is defined to be the product of the individual likelihoods

$$\frac{f(x_1)}{g(x_1)}\frac{f(x_2)}{g(x_2)}\cdot\frac{f(x_n)}{g(x_n)}$$

Let all possible samples of size n be divided into two sets with one set (the acceptance region) containing those for which the likelihood is larger than some number b and with the other set (the critical region) containing those samples for which the likelihood is less than b. It can be proved mathematically that this is the best critical region for testing the hypothesis in the sense that it minimizes the type II error for given type I error. The likelihood criterion therefore furnishes a procedure for constructing specific tests of hypotheses Other approaches to the testing problem, such as Bayes' method or the minimax principle of game theory, lead to the same criterion

If, for example, one applies the likelihood criterion to the problem of testing whether a normal population has a given mean, perhaps 10, and specifies that the type I error shall be 0.01, then the following procedure results Draw a random sample of size n, construct an 0.99 confidence interval for the population mean μ, if the number 10 lies within the interval, accept the hypothesis; otherwise, reject it

An important class of hypotheses has to do with tests of independence in multivariate populations As an example, one may consider the population of registered voters in the United States as a bivariate population with one variable being their opinions (yes, no, undecided) on some political proposition of the moment, and the other variable being geographical location. Elements of the population may be classified into a so-called contingency table as shown in Fig. 7 The hypothesis of independence asserts that the division of political opinion is unaffected by location. To test that hypothesis, a random sample of the population may be interviewed and classified into the twelve categories or cells of the contingency table. Let n_{ij} be the number of individuals falling in the cell in the ith row and

parametric methods which are valid whatever form the population distribution might take.

... techniques use the so-called or-

... le observa-

... magnitude,

... observation,

..., being the

largest A basic theorem states that on the average the sample divides the population into $n + 1$ equal parts, that is, that $1/(n + 1)$ of the population lies between any two successive order statistics Many nonparametric methods rest on this fact and its consequences.

As a simple illustration of a nonparametric estimate and confidence interval one may consider the estimation of the population median, M, given an ordered sample of size 5 consisting of x_1, x_2, x_3, x_4, and x_5 The estimate of M among the observations is simply the central sample observation x_3. The extreme observations x_1 and x_5 provide limits for a confidence interval for M, the probability level for the interval is calculated as follows One-half of the population lies to the right of M, hence the probability is 1/2 that a randomly drawn observation will lie to its right. The probability is $(1/2)^5$ that all five observations will lie to its right Similarly the probability is $(1/2)^5$ that all five will lie to its left In all other cases the sample will have at least one observation on each side of M; therefore probability that

$$(x_1 < M < x_5) = 1 - (1/2)^5 - (1/2)^5 = 30/32$$

which is about 0.94. *See* ANALYSIS OF VARIANCE; BIOMETRICS, DISTRIBUTION (PROBABILITY), EXPERIMENT, QUALITY CONTROL [A M M]

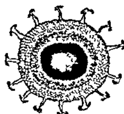
Bibliography C A Bennett and N L Franklin, *Statistical Analysis in Chemistry and the Chemical Industry*, 1954, W G Cochran, *Sampling Techniques*, 1953, W J Dixon and F J Massey *Introduction to Statistical Analysis*, 2d ed., 1957; D J. Finney *Experimental Design and its Statistical Basis*, 1955, M G Kendall, *The Advanced Theory of Statistics*, vol 1, 6th ed, 1958, vol. 2, 3d ed, 1951, A M Mood, *Introduction to the Theory of Statistics*, 1950

Statoblasts

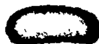
Chitin-encapsuled seedlike bodies of several types which occur in the Phylactolaemata. They are classified as sessoblasts, piptoblasts, floatoblasts and spinoblasts They are 0.26–1 5 mm long and produced in great numbers from spring to autumn (See the illustration at the top of the page)

Sessoblasts permanently attach to zooechial tubes or the substratum Floatoblasts and spinoblasts have a float of "air" cells and are, therefore, free Piptoblasts are free but have no float.

Statoblasts tide species over adverse ecological conditions such as drying or freezing which kill the colony but do not harm statoblasts. Statoblasts germinate during the season when they are produced or the following spring, but some *Lophopodella* statoblasts germinated after 50 months of drying *See* BRYOZOA [M O B O]







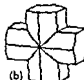


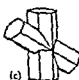
Statoblasts. *(a)* Spinoblast of *Pectinatella magnifica*. (b) Floatoblast of *Plumatella repens*. *(c)* Piptoblast or sessoblast of *Fredericella sultana*. *(d)* Sessoblast of *Stolella indica*

Staurolite

A nesosilicate mineral, $FeAl_4(SiO_4)_2(OH)_2$, that crystallizes in the orthorhombic *system* It is *frequently* in crystals, usually a combination of the vertical *prism* with the basal and side pinacoid Equally common are two types of *cruciform* penetration twins. In one type the two individuals cross at approximately 90°, in the other type they cross at about 60° The hardness is 7–7½ on Mohs scale; specific gravity is 3.7. The luster is resinous to vitreous but may be dull when the *mineral* is impure or altered, the color is reddish brown to black. *Staurolite* is a metamorphic *mineral* found in schists associated with garnet, kyanite, and tourmaline At St. Gothard, Switzerland, it is found on kyanite in parallel *orientation* In the United States it is found in schists in many states, notably New Hampshire, Massachusetts, North Carolina, Georgia, Virginia, and New Mexico. *See* SILICATE MINERALS. [C.S HU.]







Staurolite crystals (a) Simple (b) 90° and (c) 60° penetration twins (From C. S. Hurlbut, Jr., *Dana's Manual of Mineralogy*, 16th ed., Wiley, 1952)

Stauromedusae

An order of the class Scyphozoa, usually found in circumpolar regions *Haliclystus auricula* is typical The egg develops into a planula which can only creep since it lacks cilia. The planula changes into a polyp that metamorphoses directly into a combined polyp and medusa form. The medusa is composed of a cuplike bell called a ... part) and ...

dice, for example) so that minor factors not taken into account in the design will not bias the results. If, for example, learning is an important factor in operating the machines, then the experiment must be three-factor and a more elaborate design is required.

In the resulting experimental observations the effects of men and of machines are entangled, but there exists a computational technique for well-designed experiments known as the analysis of variance which will disentangle them. This enables the total variation to be broken down into parts, a variance attributable to machines, one to men, and one to interactions; if the experiment were duplicated, one could also break out a variance corresponding to experimental error. One can estimate the individual main effects of machines, the main effects of men, and individual interaction effects. In experiments involving more factors, one may be able (depending on the experimental design selected) to estimate higher-order interactions such as the second-order interactions between the main effects of one factor and the first-order interactions of two other factors.

A common hypothesis in an analysis of variance is the null hypothesis that the variance associated with some factor or interaction is not larger than the experimental error variance, the test criterion is essentially the ratio of the two variances and has the so-called F distribution when the null hypothesis is true. Rejection of the null hypothesis implies that the factor or interaction in question had a significant effect on the subject of the experiment.

Regression and correlation. The regression problem is that of estimating certain unknown constants or parameters occurring in a function which relates several variables; the variables may be random or not. By far the most easily handled cases are those in which the function is linear in the unknown parameters, and it is worth considerable effort to transform the function to that form if at all possible.

The eventual adult height, H, of a 5-year-old boy may be quite well predicted by a linear function of three variables: his own present height, B, his father's height, F, and his mother's height, M. The linear function is

$$H = a + bB + cF + dM$$

where a, b, c, and d are the unknown parameters. The three variables B, F, and M are called independent variables, the variable H is called the dependent variable, and the parameters b, c, and d are often referred to as regression coefficients.

To estimate the parameters one might draw a random sample of 5-year-old boys, measure their heights and those of their parents, then some years later measure their adult heights. One would then have sufficient data from which the parameters could be estimated by procedures entirely analogous to the methods for estimating location and scale parameters described above. In practice, one would not take so long to get the data but would use men whose childhood records were available. This method involves subtle sampling problems however; for example, persons whose records are available may have been better cared for on the average and hence taller on the average.

The data which supply the estimates of regression coefficients can also be used to estimate the standard deviation, or standard error of regression, σ. Using that estimate and a table of the t distribution, one can compute a prediction interval (analogous to a confidence interval) which will have the desired probability of including the correct adult height of a given boy.

Tall fathers sometimes have short sons and short fathers sometimes have tall sons, but generally a father's height is a good indicator of his son's height; that is, fathers' heights and their sons' heights are positively correlated. To employ a different example: The price of a commodity in a free economy is negatively correlated with the supply of that commodity, there is not a fixed relation between price and supply, but as a general proposition one goes up when the other goes down.

Statisticians have developed measures of the degree of such imprecise relationships called coefficients of correlation. The most widely used one is the Pearson or product moment correlation which is generally denoted by ρ. It measures the degree of linear association or correlation between two random variables of a bivariate or multivariate population and is defined as the mean over all the population of the product of the deviations of the two variables from their means divided by the product of their standard deviations. Mathematically,

$$\rho = \int (x - \mu)(y - \nu) f(x,y)\, dx\, dy / \sigma_x \sigma_y$$

where x and y are the random variables, μ and ν their means, σ_x and σ_y their standard deviations, and $f(x,y)$ their frequency distribution.

If the two variables are completely unrelated, then $\rho = 0$, if they have a fixed linear relation so that one can be calculated directly from the other, then $\rho = +1$ or $\rho = -1$ depending upon whether their relation is direct or reverse. Otherwise ρ will be some fraction between -1 and $+1$ with the fraction being near zero if there is poor correlation between them.

When the independent variables of a regression function are random variables, there is an equivalence between correlation coefficients and regression coefficients, after one set has been defined, the definition of the other follows automatically; mathematical formulas connecting them may be found in any statistics textbook.

Nonparametric inference. Most techniques of statistical inference rely on the central limit theorem or on sampling distributions derived from normal populations. They are practically always valid for large samples and are often valid for samples of intermediate size. However there are occasions when one cannot rely on these techniques, particularly when samples are small or when some evident peculiarity of the population (such as marked asymmetry) makes ordinarily used sampling distributions suspect. In such instances one uses non-

of a perfect gas, and wet steam is a mixture. Therefore, the properties of steam near its vaporization temperature are determined experimentally.

Data can be presented in tabular form, giving pressure, volume, entropy, enthalpy, and temperature for saturated liquid (water at the boiling point), saturated steam, and steam vapor at various temperatures of superheat. Alternatively, thermodynamic properties of steam can be presented diagrammatically. For analysis of thermodynamic cycles, the temperature-entropy chart is widely used (Fig. 2). Area on this chart is proportional to heat energy. At the critical point, steam condenses directly into water without releasing energy. In the uncharted area the pressure is higher than is usually encountered in commercial practice.

For engineering applications, the Mollier diagram presents steam data in a convenient form (Fig. 3). In the wet region, lines of constant temperature and of constant pressure are necessarily straight and coincide with *each other*. Total enthalpies above 32°F are plotted as ordinates and total entropies as abscissas.

Viscosity of steam increases with temperature and pressure. Saturated steam at atmospheric pressure has a viscosity of about 2.6×10^{-7} (lb)(sec)/(ft²). At 1000°F and atmospheric pressure it is about 6×10^{-7} and at 1000°F and 2000 lb/in.² pressure it is 13.4×10^{-7}. Similarly, the heat conductivity of steam increases with temperature and pressure. The thermal conductivities for the same three conditions for which viscosities are given are about 14, 37, and 109 Btu/(hr)(ft²)(°F)(ft).

Application. Chiefly because of its availability, but also because of its nontoxicity, steam is widely used as the working medium in thermodynamic processes. It has a uniquely high latent heat of vaporization: 1049 Btu/lb at 1 in. Hg abs and 79°F, 970 at 14.7 psia and 212°F, and 889 at 100 psia and 328°F. Steam has a specific heat in the vicinity of half that of water. For comparison, its specific heat is about twice that of air and comparable to that of ammonia. Except for a few gases such as hydrogen with a specific heat seven times that of steam, the specific heat of steam is relatively high so that it can carry more thermal energy at practical temperatures than can other usable gases. See ENTROPY; THERMODYNAMIC CYCLE; WATER.

[W.T.]

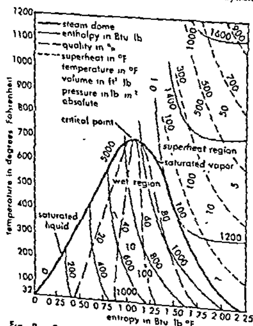


Fig. 2 Temperature-entropy chart for steam. (From J. H. Keenan and F. G. Keyes, Thermodynamic Properties of Steam, Wiley, 1936)

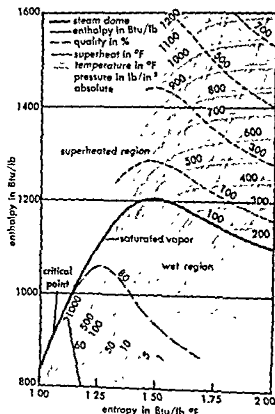
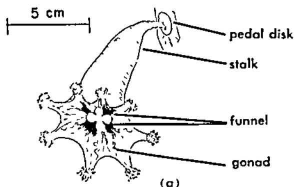


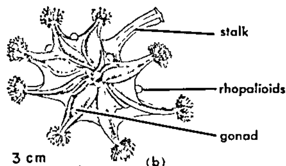
Fig. 3 Mollier or enthalpy-entropy diagram for steam. (From R. Mollier, Mollier's Steam Tables and Diagrams, Pitman, 1927)

Steam boiler

A pressurized system in which water is vaporized to steam as a desired e... from a s... products ... thus gene... ...y be used directly as a heating medium, or as the working fluid in a prime mover to convert thermal energy to mechanical work, which in turn may be converted to electrical energy. Although other fluids are sometimes used for these purposes, water is by far the most common because of its economy and suitable thermodynamic characteristics.

The physical sizes of boilers range from small domestic units to large commercial stationary units, comparable to a multi-story building, 150 ft





Stauromedusae (a) Lucernaria. (b) Haliclystus. (From L. H. Hyman, The Invertebrates, vol 1, McGraw-Hill, 1940)

groups of short, capped tentacles and eight sensory bodies, called anchors, on its margin The mouth, situated at the center of the calyx, has four thin lips and leads to the stomach in which gastral filaments are arranged in a row on either side of each interradius. Though sessile, the medusa can move in a leechlike fashion by alternate attachment and release of the pedal disk, using the substratum as an anchor. *See* SCYPHOZOA. [T.U.]

Steam

Water vapor or water in its gaseous state Steam is by far the most used working medium in external combustion engines such as steam turbines; it is also widely used as a heating medium

Steam conditions. The temperature at which steam forms depends entirely on the pressure in the boiler or steam generator (Fig 1). The steam formed in a boiler is in temperature equilibrium with the water. Under this condition, with steam and water at the same temperature, the steam is termed saturated. The steam can be entirely vapor, that is, it can be dry, or it can carry entrained moisture in suspension and be wet. It may also be contaminated with other gases such as air, in which case it is a mixture (*see* DALTON'S LAW). After the steam is removed from contact with the water, without changing its pressure, the steam can be heated further If initially wet, the additional heat will first dry it and then raise it above its saturated (evaporation) temperature. Under this condition the steam is superheated

The heat energy in steam can be divided into three parts: (1) enthalpy of the liquid required to raise the water from its initial temperature (usually considered to be 32°F) to the boiling temperature, (2) enthalpy of vaporization required to convert the water to steam at the boiling temperature, and (3) enthalpy of superheat that raises the steam to its final temperature (*see* ENTHALPY). As the steam performs its thermodynamic function giving up its heat energy, it loses its superheat, becomes wet, and finally condenses to hot water. While wet, the steam has a quality that decreases as the per cent of dry saturated steam present in the wet steam decreases. Dry steam has a quality of 100%.

Properties as a gas. Superheated steam at temperatures well above the boiling temperature for the existing steam pressure follows closely the laws of a perfect gas. Thus, $pv \approx 85.8T$ and $k \approx 1.3$ where p is pressure in lb/ft², v is volume in ft³/lb, and T is absolute temperature in °R, and k is the ratio of specific heat at constant pressure to specific heat at constant volume. However, the behavior of dry saturated steam departs from that

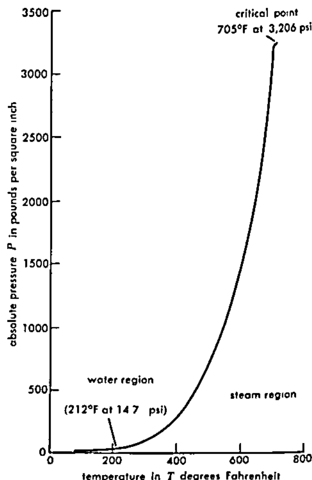


Fig 1 Boiling temperature for steam increases with pressure; above 705°F steam (water in its gaseous state) and water in its liquid state occupy the same specific volumes and therefore do not separate because of density differences alone as they do during boiling at pressures below 3206 psi.

mills The ease with which the direction of rotation of a reciprocating steam engine is reversed is a further advantage in these applications. However, because the steam turbine has a smaller size for [illegible] highest effi- [illegible] ciprocating [illegible] mover for electric generating stations (see STEAM TURBINE). The greater ease with which internal combustion engines start has resulted in their replacing the steam engine for traction service When used with a gear shift or a hydraulic torque converter, these newer type engines can also develop high starting torque. Where process steam is used, steam engine power is, however, a low-cost by-product.

Cylinder action. A typical steam reciprocating engine consists of a cylinder fitted with a piston (Fig 1) A connecting rod and crank shaft convert the piston's to-and-fro motion into rotary motion. A flywheel tends to maintain a constant output angular velocity in the presence of the cyclically changing steam pressure on the piston face. A valve admits high-pressure steam to the cylinder and allows the spent steam to escape (Fig. 2)

The power developed by an engine depends on the pressure and quantity of steam admitted per unit of time to the cylinder The pressure varies during the stroke, having an average designated as the mean effective pressure p lb/sq in The quantity of steam per minute, or the steam rate, is the volume of the cylinder filled by steam once each

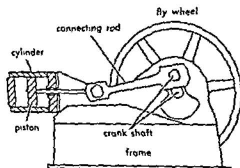


Fig 1 Principal parts of horizontal steam engine

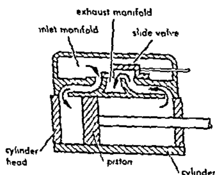


Fig 2 Single-ported slide valve on counterflow double-acting cylinder

revolution multiplied by the revolutions per minute n The volume is in turn the piston area a in sq in times the length of piston stroke l in ft. The product of these design dimensions gives the horsepower rating P for a single-acting steam engine:

$$P = plan/33{,}000$$

where 33,000 converts ft lb per minute to horsepower. With steam admitted alternately to each side of the piston, as in the double-acting engine of Fig. 2, the theoretical power developed is approximately twice this value. The power developed by an actual engine is 70–90% of the ideal power because of friction and other losses

Engine types. Engines are classified as single- or double-acting, and as horizontal, as in Fig 1, or vertical depending on the direction of piston motion. If the steam does not fully expand in one cylinder, it can be exhausted into a second larger cylinder to expand further and give up a greater part of its initial energy. Thus, an engine can be compounded for double or triple expansion In counterflow engines, steam enters and leaves at the same end of the cylinder; in uniflow engines, steam enters at the end of the cylinder and exhausts at the middle

Steam engines can also be classed by functions, engines for different services being built to optimize the characteristics most desired in each application Stationary engines drive electric generators, in which constant speed is important, or pumps and compressors in which constant torque is important. Governors acting through the valves hold the desired characteristic constant. Marine engines require a high order of safety and dependability

Valves. The extent to which an actual steam piston engine approaches the performance of an ideal engine depends largely on the effectiveness of its valves. The valves alternately admit steam to the cylinder, seal the cylinder while the steam expands against the piston, and exhaust steam from the cylinder (see CARNOT CYCLE; THERMODYNAMIC CYCLE). The many forms of valves can be grouped as sliding valves and lifting valves (Fig. 3)

Sliding valves are ones such as the D valves (Fig 2). They may be combined with expansion valves. A common sliding valve is the rocking Corliss valve; it is driven from an eccentric on the main shaft like other valves but has separate rods for each valve on the engine. After a Corliss valve is opened, a latch automatically disengages the rod and a separate dashpot abruptly closes the valve Exhaust valves are closed by the rods as with other sliding valves.

Lifting valves are more suitable for use with high-temperature steam They, too, are of numerous forms, the poppet valve being representative

The valves are driven through a crank or eccentric on the main crankshaft The crank angle is set to open the steam port near dead center, where the piston is at its extreme position in the cylinder

high by 80 ft wide and 80 ft deep The larger units are assembled at the permanent site. In terms of steam generating capacities, commercial boilers range from a few hundred pounds of steam per hour to more than 4,000,000 lb/hour.

The pressure at which steam is generated extends from a few pounds per square inch above atmospheric pressure, to several thousand pounds per square inch, depending upon requirements of the process served Pressure-part components must be strong enough to withstand the generated steam pressure, and must be maintained at acceptable temperatures, by transfer of heat to the fluid, to prevent loss of strength of the construction materials by overheating or destructive oxidation

Being in the class of durable goods, boilers that receive proper care in operation and maintenance function satisfactorily for several decades Thus the types of boilers found in service at any time represent a wide span in the stages of development in boiler technology.

The earliest boilers, used at the beginning of the industrial era, were simple vats or cylindrical vessels, made of iron or copper plates riveted together and supported over a furnace fired by wood or coal. Connections were made for offtake of steam and for the replenishment of water Evolution in design for higher pressures and capacities led to the use of steel, and to the employment of tubular members in the construction, to increase the amount of heat-transferring surface, at first for the passage of hot gases through tubes submerged in the water space of the vessel (*see* FIRE-TUBE BOILER), and later by arrangements of multiple tubes containing the water (*see* WATER-TUBE BOILER), which were exposed on their outer surface to contact with hot gases.

The over-all functioning of steam generating equipment is governed by thermodynamic properties of the working fluid (*see* STEAM) By the simple addition of heat to water in a closed vessel, vapor is formed which has greater specific volume than the liquid, and can develop increase of pressure to the

can be maintained at any desired value, and thus be held within the limits of safety of the construction

Addition of heat to steam, after its generation, is accompanied by increase of temperature above the saturation value (*see* SUPERHEATER). The higher heat content, or enthalpy, of superheated steam permits it to develop a higher percentage of useful work by expansion through the prime mover, with a resultant gain in efficiency of the power generating cycle.

If the steam generating system is maintained at pressures above the critical, by means of a high-pressure feed-water pump, water is converted to a vapor phase of high density equal to that of the water, without the formation of bubbles. Further heat addition causes superheating, with corresponding increase in temperature and enthalpy. The most advanced developments in steam generating equipment have led to units operating above critical pressure, at 4500–5000 psi.

Superheated steam temperature has advanced through the years from about 500°F to the present practical limits between 1050 and 1100°F. Progress in boiler design and performance has been governed by the continuing development of improved materials for superheater construction having adequate strength and resistance to oxidation for service at elevated temperatures. For the high temperature ranges, complex alloy steels are used in parts of the assembly. *See* STEAM GENERATING UNIT.

[F.G.F.]

Steam condenser

A heat transfer device used to condense steam to water by removing the latent heat and absorbing it in a heat receiving fluid, usually water.

Steam condensers may be classified as contact or surface. Condensing takes place in the contact condenser in a chamber where the steam and cooling water mix. In the surface condenser the condensing process takes place separated from the cooling water by a metal wall, which forms the condensing surface

Both contact and surface condensers are used in process systems and for serving engines and turbines for generating power. Modern practice has confined the use of contact condensers, for the most part, to such process systems as those involving vacuum pans, evaporators, or dryers. The steam surface condenser is used chiefly in power generation but is also used in process systems, especially those in which condensate recovery is important

Surface condensers used for power generation are of the high-vacuum type Because their main purpose is to effect a back pressure at the turbine exhaust, for economical station heat rate and fixed cost, they must be designed for high heat transfer rates, minimum steam-side pressure loss, and efficient air removal

Condenser sizes have increased with the size of turbine generators The largest, in a single shell, is approximately 200,000 ft^2, the condensing steam space occupies a volume of about 30,000 ft^3 Larger units have been built with the surface divided equally between two shells. Power plant condensers require 70 100 lb of cooling water to condense 1 lb of steam Normally 0.5 ft^2 of surface is required for each kilowatt of generating capacity *See* CONDENSER, VAPOR

[J F S]

Steam engine

A machine for converting the heat energy in steam to mechanical energy. Compared to other engines used as prime movers, the reciprocating steam engine, with which this article deals, develops its full rated torque at any speed from rest to full throttle This variable-speed, high torque characteristic is desirable in traction engines such as locomotives, in hoists, and in heavy machinery such as rolling

Fig. 4 Events during one cycle of piston operation (a) In ideal engine (b) As depicted on the indicator card of a noncondensing steam engine.

multistage engines, the steam can be reheated between stages. Engine sizes are generally limited in current practice to 1000 hp, 700 ft/min piston speed, 200 psi, and 600°F [W.T.]

Steam generating unit

The complete assembly of equipment in a modern steam power plant which operates as a unit to produce high-pressure, high temperature steam.

Essential steps of the steam generating process comprise making heat available by combustion of fuel, and absorbing this heat (1) to raise the feed water to boiling temperature, (2) to evaporate the water into steam, and (3) to increase the enthalpy of the steam by heating it above saturation temperature. For improvement of over-all thermal efficiency air heaters recover some of the remaining heat from the cooled products of combustion.

For this aggregation of functions, the traditional name of boiler is inadequate and has been superseded by the broader designation of steam generating unit. A complete unit includes fuel-burning equipment, *furnace*, and all components that absorb heat from the products of combustion, as well as major auxiliary equipment such as blowers and pumps required for operation (Fig. 1).

Water-cooled furnaces are included as ... [illegible]

surface required to form the furnace walls is frequently sufficient to accomplish all of the evaporating function in cases where much of the available heat is needed for superheating to high steam temperatures. For low superheat service, a generating tube bank is used to supplement the furnace walls. Proportioning and arrangement of the various classes of heating surface may be greatly diversified in different designs to meet the requirements of fuel characteristics and specified final steam pressure and temperature conditions.

In most units, water recirculates in the generating sections after separation of steam, the flow being produced by natural circulation or, in some designs, being assisted by a pump.

A different class of unit, which shows increasing prominence, is the forced flow once-through steam generator (Fig. 2). It has no separating drum. Feed water enters the system, passes through heated continuous circuits to the outlet, being converted to steam and superheated in transit. The degree of superheat is regulated by rate of firing. This type of system is mandatory for units operating above critical pressure because steam and water of equal densities will not separate. The system requires feed water of exceptional purity to avoid formation of deposits in the heated circuits or transport of solids to the turbine. Such water conditions are now commercially attainable. Boilers of this type are practical for pressures above or below critical and provide advantages in over-all economy of power generation.

Fig. 1 Steam generating unit.

The angle between valve crank and connecting rod crank is slightly greater than 90°, the excess being the angle of advance.

So that the valves will open and close quickly, they are driven at high velocity with consequently greater travel than is necessary to open and close the ports The additional travel of a sliding valve is the steam lap and the exhaust lap. The greater the lap, the greater the angle of advance to obtain the proper timing of the valve action.

Engine power is usually controlled by varying the period during which steam is admitted. A shifting eccentric accomplishes this function or, in releasing Corliss and in poppet valves, the eccentric is fixed and cutoff is controlled through a governor to the kickoff cams or latch that allows the valves to be closed by their dashpots

For high engine efficiency, the ratio of cylinder volume after expansion to volume before expansion should be high. The volume before expansion into which the steam is admitted is the volumetric clearance. It may be determined by valve design and other structural features. For this reason, valves and ports are located so as not to necessitate excessive volumetric clearance.

Indicator card. The combined action of valves and piston is studied by means of an indicator card A card is mounted on a small drum. The drum is rotated back and forth by the piston rod through a reducing mechanism. The steam pressure in the cylinder actuates a pencil or stylus to move up and down along the card The resulting diagram is a record of cylinder steam pressure as a function of piston position. The events of a cycle of engine operation can be identified on the diagram (Fig. 4).

The abscissa, controlled by piston motion, is proportional to steam volume, and the ordinate is proportional to steam pressure. The area enclosed by the diagram is proportional to the work exerted by the steam on the piston. The mean effective pressure is that pressure which, if exerted during the full stroke, would produce the same total work as the actual varying pressure From a knowledge of the scale of the indicator card, the value of p in the equation for engine horsepower can be determined

Efficiency of the engine depends on the decrease in energy in the steam between inlet and outlet (see ENTHALPY) It also depends on thermal losses, diagram losses, and mechanical losses. Thermal losses are those due to initial condensation, in which part of the incoming steam condenses on the cylinder walls, and to radiation, in which the heated cylinder walls give off heat to their external surroundings.

Diagram losses arise from incomplete expansion of the steam and from throttling during admission and exhaust. Mechanical losses are produced by the friction effects of piston in cylinder, packings, bearings, valves, and valve gear. Thermal losses are minimized by multistage expansion, superheated steam, and uniflow cylinder arrangements. With

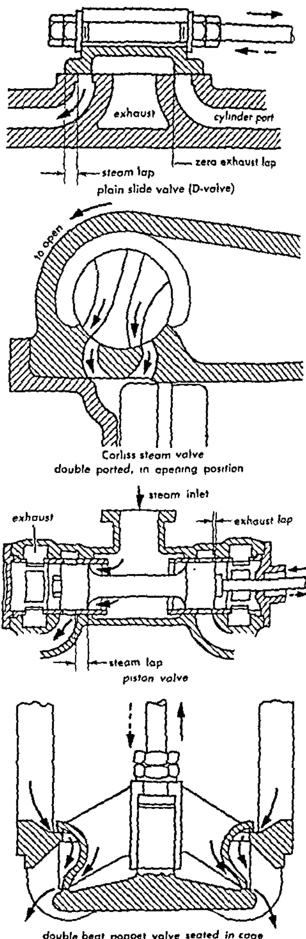


Fig 3 Typical steam engine valves in closed positions Arrows show path steam will travel when valves (shaded part) open.

driven pump controlled from the boiler water level then forces the condensate back to the boiler.

In large buildings extending over a considerable area, it is difficult to locate all heat exchangers above the boiler water level or return piping. For these systems a vacuum pump is used that maintains a suction below atmosphere up to 25 in. of mercury in the return piping thus creating a positive return flow of air and condensate back to the pumping unit. Subatmospheric systems are similar to vacuum systems, but in contrast provide a means of partial vacuum control on both the supply and return piping so that the steam temperature can be regulated to vary the heat emission from the heat exchanger in direct proportion to the heat loss from the structure. See COMFORT CONTROL. [J W J]

Steam jet ejector

A steam-actuated device for pumping compressible fluids, usually from subatmospheric suction pressure to atmospheric discharge pressure. A steam jet ejector is most frequently used for maintaining vacuum in process equipment in which evaporation or condensation takes place. Because of its simplicity, compactness, reliability, and generally low first cost, it is often preferred to a mechanical vacuum pump for removing air from condensers serving steam turbines, especially for marine service.

Principle. Compression of the pumped fluid is accomplished in one or more stages, depending upon the total compression required. Each stage consists of a converging-diverging *steam nozzle*, a suction chamber, and a venturi shaped diffuser. Steam is supplied to the nozzle at pressures in the range of 100 to 250 psig. A portion of the enthalpy of the steam is converted to kinetic energy by expanding it through the nozzle at substantially constant entropy to ejector suction pressure where it reaches velocities of from 3000 to 4500 ft/sec. The air or gas with its vapor of saturation, which is to be pumped and compressed is entrained, primarily by friction in the high-velocity steam jet. The impulse of the steam produces a change in the momentum of the air or gas vapor mixture as it mixes with the motive steam and travels into the converging section of the diffuser. In the throat or most restricted area of the diffuser, the energy transfer is completed and the final mixture of gases and vapor enters the diverging section of the diffuser, where its velocity is progressively reduced. Here a portion of the kinetic energy of the mixture is re-converted to ... in enthalpy ... higher pre... ejector (se... ...ESTER). The compression ratios selected for each stage of a steam jet ejector usually vary from about 1 to 7.

Application. Two or more stages may be arranged in series depending upon the total compression ratio required (Fig. 1). Two or more sets of series stages may be arranged in parallel to accommodate variations in capacity.

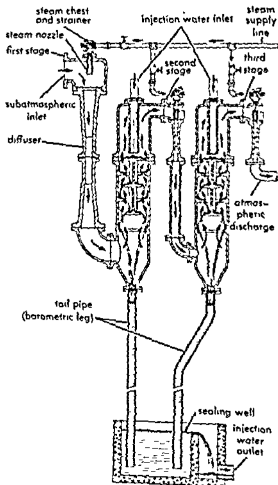


Fig. 1. *Typical multistage steam jet ejector with contact barometric condensers, first and second stage condensing, third stage noncondensing.*

Vapor condensers are usually interposed between the compression stages of multistage steam jet ejectors to condense and remove a significant portion of the motive steam and other condensable vapors (Fig. 2). This action reduces the amount of fluid to be compressed by the next higher stage and results in a ... Both su... used fo...

Eject... used as air pumps for steam condensers that serve turbines are usually two stage and are equipped with inter- and after condensers of the surface type. The steam condensed is drained through traps to the main condenser and returned to the boiler feed system. Ejectors used as vacuum pumps in process systems may be equipped with either surface condensers or contact condensers of the barometric type between or after stages or both. They may be single or multistage machines. High vacuum process ejectors with as ... seven stag... pro... cha... bec... manufactured economically from almost any corrosion resistant material.

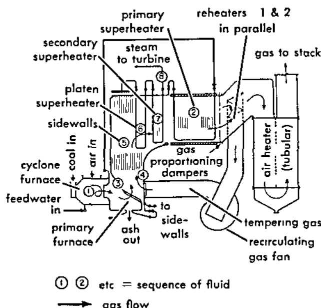


Fig 2. Once-through boiler.

A modification of the once-through circuit, applicable to units designed for operation below the critical pressure, employs a small centrifugal separating drum between the generating and superheating sections. The separated water, containing solids that may have entered with the feed water, is discharged from the system through a heat exchanger In this arrangement, control of final steam temperature is conventional. *See* Boiler feed water, Boiler water; Economizer and air heater, Feed water; Fire-tube boiler; Marine boiler, Raw water; Reheating; Steam boiler; Steam separator; Steam temperature control, Superheater, Water-tube boiler [F G E.]

Steam heating

A heating system that uses steam generated from a boiler. The steam-heating system conveys steam through pipes to heat exchangers, such as radiators, convectors, baseboard units, radiant panels, or fan-driven heaters, and returns the resulting condensed water to the boiler. Such systems normally operate at pressure not exceeding 15 pounds per square inch gage (psig) and in many designs the condensed steam returns to the boiler by gravity due to the static head of water in the return piping With utilization of available operating and safety control devices, these systems can be designed to operate automatically and safely with minimum maintenance and attention.

One-pipe system. In a one-pipe steam heating system, a single main serves the dual purpose of supplying steam to the heat exchanger and conveying condensate from it (Fig. 1). Ordinarily, there is but one connection to the radiator or heat exchanger, and this connection serves as both the supply and return; separate supply and return connections are sometimes used. Because steam cannot flow through the piping or into the heat exchanger until all the air is expelled, it is important to provide automatic air-venting valves on all exchangers and at the ends of all mains These valves may be of a type which closes whenever steam or water comes in contact with the operating element but which also permits air to flow back into the system as the pressure drops A vacuum valve closes against subatmospheric pressure to prevent the return of air.

Two-pipe system. A two-pipe system is provided with two connections from each heat exchanger, and in this system steam and condensate flow in separate mains and branches (Fig 2). A vapor two-pipe system operates at a few ounces above atmospheric pressure, and in this system a thermostatic trap is located at the discharge connection from the heat exchanger which prevents steam passage, but permits air and condensation to flow into the return piping.

When the steam condensate cannot be returned by gravity to the boiler in a two-pipe system, an alternating return lifting trap, condensate return pump, or vacuum return pump must be used to force the condensate back into the boiler. In a condensate return-pump arrangement, the return piping is arranged for the water to flow by gravity into a collecting receiver or tank, which may be located below the steam-boiler water line. A motor-

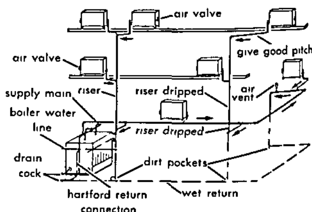


Fig 1 Preferred up-feed gravity one-pipe air-vent system (*American Society of Heating and Air-Conditioning Engineers, Inc*)

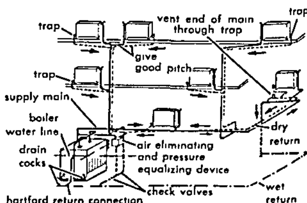
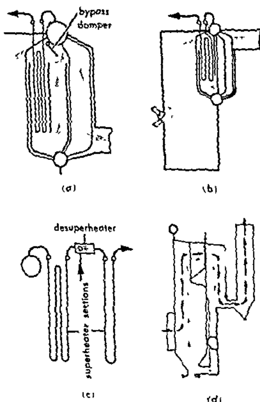


Fig 2 Typical up-feed two-pipe system with automatic return trap (*American Society of Heating and Air-Conditioning Engineers, Inc*)

Steam temperature control

Adjustments to a steam generating unit to produce steam at the required temperature. Among the factors that affect the performance of a steam generating unit are ash or slag deposits on the heating surface, changes in the proportioning of fuel and combustion air, or changes in feed-water temperature. Abnormally low steam temperature results in lowered efficiency of the power generating cycle, while limits to high temperature are imposed by the strength and durability of materials used in superheater construction. Steam temperature control is therefore a matter of primary concern in the design of complex modern units.



(a-d) Principal means for controlling steam temperature

Reg [illegible]

[illegible] around the superheater by means of adjustable dampers; (b) selective use, or altered positioning of burners, to change the location of the combustion zone; (c) controlled attemperation or desuperheating of steam between the primary and final stages of the superheater, by injection of water spray, or passage of a portion of the steam through a heat exchanger submerged in the boiler water and (d) mixing of recirculated cooled flue gases with the gas stream ahead of the superheater. These methods of control may be adjusted manually or automatically. See FURNACE (STEAM GENERATING); STEAM GENERATING UNIT [F.C.E.]

Steam turbine

By far the most widely used and greatest power-producing turbines are those driven by steam. The turbines convert the energy in the steam into rotating energy of an output shaft. The steam can be generated in boilers using a wide variety of fuels in as large quantities as may be desired to produce the required power, and the power produced is not dependent upon the availability of large quantities of water with an available head as is required for a water turbine, or a strong and steady wind as for a windmill. Steam turbines aggregating at least 10,000,000 horsepower capacity are built in the world every year; they range in size from a few horsepower to several hundred thousand horsepower. Manufacturers of steam turbines are located in every industrial country.

Turbine parts. The steam turbine consists of the following essential parts (Fig. 1):

1. A casing, or shell, usually divided at the horizontal centerline, with the halves bolted together for ease of assembly and disassembly, and containing the stationary blade system.

2. A rotor, carrying the moving blades (buckets or vanes) either on wheels or drums, with bearing journals on the ends of the rotor.

3. A set of bearings attached to the casing, to support the shaft.

4. A governor and valve system for regulating the speed and power of the turbine by controlling the steam flow, and an oil system for lubrication of the bearings and, on all but the smallest machines, for operating the control valves by a relay system connected with the governor.

5. A coupling of some sort to connect with the driven machine.

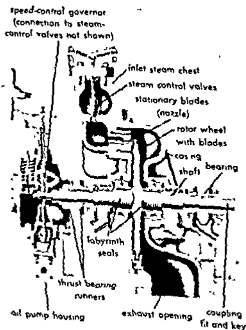


Fig. 1 Cutaway view of small, single-stage steam turbine (General Electric [illegible]

subatmospheric inlets
steam supply line
steam chest and strainer
first stage
suction chamber
steam nozzle
second stage
diffuser
safety valve
atmospheric outlet
cooling water outlet
atmos- pheric outlet
air passages
cooling water inlet
pass
drains
pass
inter condenser second vapor pass
after condenser second vapor pass

Fig 2 Typical twin-element multistage steam jet ejector with surface inter- and after-condenser, first and second stages condensing

The number of stages of compression usually used for various suction pressures with atmospheric discharge pressure is as follows:

No of ejector stages	*Range of suction pressure*
1	3–30 in. Hg abs
2	0 4–4 in Hg abs
3	1–25 mm Hg abs
4	0 15–3 mm Hg abs
5	20–300 μ
6	5–20 μ
7	1–5 μ

Ejectors are also made which use air or other gases instead of steam as the energy source [J F S]

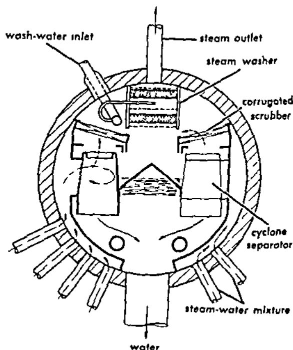
Steam separator

A device for separating the mixture of steam and water discharged into the drum by the generating circuits. In low-pressure boilers, gravitational forces are usually sufficient to produce separation and establish a water surface, which is normally maintained at the center of the drum.

In the higher ranges of pressure, however, with progressive decrease in density differential between water and steam, and with the accompanying use of thick-walled drums of small diameter, the force of gravity is insufficient to prevent the carryover of water droplets in the outgoing steam or entrainment of steam bubbles in the downcomer flow, both of which are undesirable.

Baffle arrangements are frequently installed to deflect the water away from the steam outlet. Virtually complete separation can be obtained in high-duty units by use of cyclones (as illustrated) or similar centrifugal devices, which utilize the energy of riser discharge to create separating forces several times that of gravity. Extremely small liquid droplets can be removed by passing the steam at low velocity through scrubber baffles, composed of closely spaced intermeshed corrugated plates, which present sinuous flow paths and a large surface area for liquid interception and adherence. Steam washers may also be used. *See* BOILER WATER.

[F.C F]



Means for separating water from steam

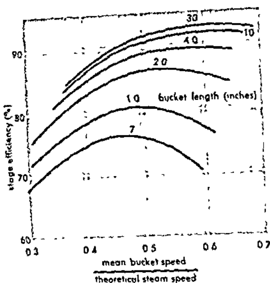


Fig. 3. Approximate efficiency of typical steam turbine stages.

at 1100°F, and at least one large unit designed for 1200°F.

The blade system is the principal part of the turbine; it is the part that extracts power from the flowing steam. The basic requirements to be satisfied by a steam turbine in producing a given amount of power from steam at given conditions are listed below.

1. The turbine must pass the necessary flow of steam so that the power output is practically possible. The required rate of steam flow is proportional to the developed power and inversely proportional to the available energy in the steam and the turbine efficiency.

2. The turbine must be provided with sufficient number and diameter of rotating blade rows so that at the desired speed of rotation, the ratio of blade speed to steam speed will yield a satisfactory efficiency (Fig. 3).

3. Each stage must be so proportioned that the steam passing through it will be at the desired pressure and will acquire the desired velocities in the stationary and moving blade rows, so that the total energy will be divided as desired.

4. The exhaust end stages must be large enough in area so that the steam may leave them without excessive velocity. As mentioned before, this requirement gives rise on large units to multiflow exhaust ends.

5. The design rotative speed must be selected; sometimes the turbine is arranged to drive through a gear, in which cases the turbine speed can be chosen independently of the speed of the driven machine, with the gear ratio being chosen to suit. On alternating-current generator drives, the speed of the generator is fixed by the number of poles in the generator and the desired frequency, and the turbine speed is usually fixed at the required generator speed except for small power units where gears may be used.

The turbine, in order to have good efficiency, must have a sufficient number of stages of a large enough diameter to accommodate the total energy allotted to the machine with a velocity ratio sufficient to achieve the efficiency. It will be evident that this objective can be attained with small-diameter wheels at high rotative speeds, and, conversely, requires larger wheels at lower speeds. Since in general the amount of material in the turbine varies about as the cube of its diameter, the smallest, lightest-weight machine will generally be the one with the highest rotative speed.

Application requirements. Steam turbines are used for a very wide variety of power drives. The very largest, 500,000 kw, as well as smaller machines, ranging down in size to 1000 kw, are used to drive electric generators for electric power. In sizes up to 70,000 hp steam turbines are used for driving ships' propellers. The drive is through reduction gearing since the turbine speed is always higher than desirable for a marine propeller. They are used for practically all other types of mechanical drives as well, including pumps, blowers, air and other gas compressors, and paper machines. A useful feature in many of these applications is the fact that the turbine can be equipped with an adjustable-speed governor, and thus be made capable of producing power over a wide range of rotative speed. Its efficiency varies with speed (Fig. 3).

It is also easily possible, and in some applications very useful, to extract steam from the turbine intermediate between its inlet and its exhaust, for heating purposes or for process work. The turbine valve system can be made to control the pressure (and therefore the temperature) of this extracted steam quite closely, which is valuable in controlling chemical processes and in paper making. The turbine can be made to exhaust at a desired pressure to process equipment instead of to the atmosphere or to a condenser.

Large turbines for electric power production are designed for the efficient use of steam in a heat cycle that involves extraction of steam for feed-water heating, resuperheating of the main steam flow, and exhausting at the lowest possible pressure consistent with the temperature of the available condenser cooling water. Such machines are built with more than one casing, and usually with more than one exhaust stage (Fig. 2).

Machine requirements. In addition to these broad requirements, there are several basic characteristics that determine the general arrangement of a steam turbine.

Centrifugal stresses. Because the steam turbine is essentially a high-speed machine, the rotating parts are of necessity designed with centrifugal stress very much in mind, although in some cases this may not be limiting. This consideration gives rise to radially tapered blades, and various designs of strong attachments for rotor blades to rotors. The most difficult problems of this kind are associated with the longest blades in the exhaust end,

6 Pipe connections to a supply of steam at the inlet, and to an exhaust system at the outlet of the casing or shell.

Turbine types. Steam turbines are classified in various ways: (1) by mechanical arrangement, as single casing, cross compound (more than one shaft side by side), tandem compound (more than one casing with a single shaft); (2) by steam flow direction—nearly all turbines are built for axial flow, though there are some with radial flow; (3) by steam cycle, that is, condensing, noncondensing, automatic extraction, reheat; (4) by number of exhaust flows of a condensing unit, as single, double, triple flow, and so on. Units with as many as six exhaust flows have been built. Often a machine will be described by several of these terms.

The simplest type of steam turbine has one stage; that is, one row each of stationary and of moving blades (Fig. 1). Such turbines are commonly used for power outputs of a few hundred horsepower at most, with moderate inlet pressures and temperatures, and for atmospheric or higher pressure at the exhaust. Under these conditions it is possible to use the steam with adequate efficiency in a single stage.

For large power output, and for the high inlet pressures and temperatures and low exhaust pressures which are required for good thermal efficiency, a single stage is not adequate. Steam under such conditions has high available energy and for its efficient utilization the turbine must have many stages in series, each taking its share of the total energy and contributing its share of the total output. Also, under these conditions the exhaust volume flow becomes large, and it is necessary to have more than one exhaust stage to avoid a high leaving velocity and consequent high kinetic energy loss; for example, a large turbine may have three exhaust stages in parallel (Fig. 2). Between the designs shown in Figs. 1 and 2 are many sizes and arrangements to meet as many operating requirements.

Steam requirements. Steam inlet pressures vary widely. Up to 1000 hp, inlet pressures of 100–400 lb per sq in. gage are common; from 1000 to 10,000 hp, 300–800 lb; from 10,000 hp to 100,000 kw, 600–2000 lb, above 100,000 kw, 1200–3500 lb. A few developmental units use pressures ranging up to 5000 lb per sq in. Steam temperatures on small units range up to 800°F, and on the larger machines up to 1050°F, with again a few

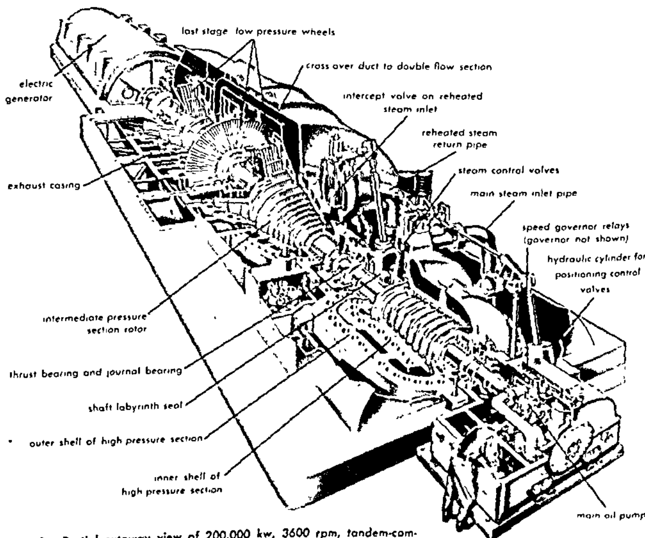


Fig. 2. Partial cutaway view of 200,000 kw, 3600 rpm, tandem-compound, triple-flow reheat steam turbine with direct connected electric generator. (General Electric Co.)

but the smallest machines, the governor signal must be power-relayed to furnish large enough forces to position the steam control valves accurately, closing them when the speed increases and opening them when the speed decreases. The speed-limiting governor, or overspeed governor, is so made as to go through its full stroke immediately if the turbine speed reaches a predetermined value above normal speed, usually about 10%. It is made to actuate a quick-closing emergency stop valve

These two kinds of governors and their separate valve systems offer two lines of defense against a possible disastrous overspeeding should the turbine suddenly lose its load Steam turbines cannot usually be designed to be safe at runaway speed. The governor systems are designed with great attention to their reliability

The governors and the valves which they control are fast in action; the valves will usually travel from wide open to fully closed in a fraction of a second (*see* GOVERNOR).

Oiling system The turbine shaft runs at a high speed and its bearings must be positively and continuously supplied with oil· they will burn out quickly if the oil supply fails The integrity of the oil system must be assured two oil pumps, a main pump and a standby, driven by separate power sources are usually provided on all but the smallest machines. The oil will also usually be used as a hydraulic power fluid to operate the steam control valves It is customary to provide for closing the steam control valves to shut down the turbine if the oil supply pressure fails.

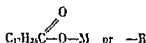
Aerodynamic design of blade system The blading design for highest efficiency, especially for the larger sizes of turbines, draws upon modern aerodynamic theory Classic forms of impulse and reaction blades merge together in the three-dimensional design required by modern concepts of loss-free fluid flow To meet the theoretical steam flow requirements, and to minimize centrifugal forces on the rotor blades and their attachments the sections change in shape along a blade (Fig 5)

Materials used in steam turbines. The materials used in the making of a steam turbine must be sound strong and of the best quality Casings or shells are almost always made of cast steel, and for the higher pressures and temperatures it is alloyed with molybdenum, vanadium, chromium and sometimes other elements for greater strength Small turbine casings for the lower range of temperatures may be of a good grade of cast iron, and the low pressure casings of large machines may also be of this material The rotors are steel forgings alloyed as may be necessary for greater strength with some of the materials mentioned above The rotor may be built up, that is made of separate wheels shrunk on a shaft, or it may be machined, wheels and all, out of one solid large forging or it may be a number of disks welded together The blades, both stationary and rotating are usually of a low-carbon 12–13% chrome steel, which is very rust resistant

[P.H K.]

Stearate

A salt (soap) or ester of stearic acid having the general formula



$$C_{17}H_{35}C(=O)—O—M \quad or \quad —R$$

and formed by replacing the carboxylic hydrogen by a metal (M) to give a salt, or by an organic radical (R) to give an ester. Stearates occur in nature chiefly as the glyceryl ester, found in substantial amounts in animal and vegetable fats The esters of long-chain alcohols are known as waxes Other esters of monohydric and polyhydric alcohols are used in cosmetics, lacquers and nonionic surface-active agents Alkali-metal salts are water soluble, and with the similar oleates and palmitates are the major components of toilet and laundry soaps. Other metal salts are used in paints, waterproofing, pharmaceuticals, cosmetics, and fungicides. *See* FAT AND OIL, EDIBLE, FAT AND OIL, NONEDIBLE; SOAP AND DETERGENT; WAX, ANIMAL AND VEGETABLE.

[E.H H.]

Steel

Ferritic steels are those having a body-centered cubic crystal lattice at room temperature, as distinguished from austenitic steels which have a face-centered cubic lattice at room temperature. Ferritic steels constitute about 99% of the steel output of the United States. Steel is the basic industrial material of the twentieth century

Known for its great strength, steel is an alloy of iron and carbon Iron itself is not particularly strong, but its abundance and the remarkable increase in strength which becomes possible when

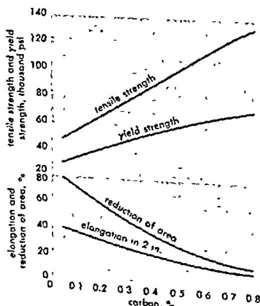


Fig 1. Effect of carbon content on the tensile properties of hot-worked carbon steels.

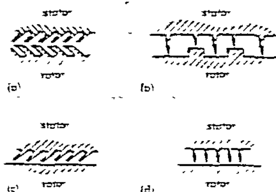


Fig. 4. (a–d) Typical labyrinth seals. Seals (a) and (c) offer most resistance when flow is from left to right.

or with the hottest blades at the inlet end. Stresses in wheels or rotor bodies are also often limiting.

Casing or shell stresses. The high pressure at the inlet must be contained in a properly strong casing or shell; these casings become quite massive, especially on large units at high steam pressures. This casing must also be split at the horizontal centerline for assembly and maintenance, and the halves bolted together to be leak-tight. The smaller the casing diameter, the easier will this be to accomplish, so for high-pressure turbines there is a definite advantage to small-diameter high-speed rotors. Sections of large turbines are often made with double shells, an inner one with high pressure inside and surrounded by an intermediate pressure which is contained by the outer shell (Fig. 2). Many large bolts are required to hold these shells together against the high inside steam pressure.

Rotor blades or buckets. The blades must be strong enough to withstand high centrifugal forces and vibration. Any blade row, and especially the rotor blades, may vibrate in one of a number of modes, each mode having a different frequency. If a mode happens to be resonant with a stimulus from the steam forces, a destructive vibration may ensue. Care must be used in design to avoid such a condition; if it is not possible to avoid all such vibrations, as is apt to be the case with a variable-speed turbine, the blades must be made strong enough to withstand the vibrations that do occur.

Shaft sealing against steam leakage. It is necessary to minimize to the greatest possible extent the leakage of steam along the shaft, both at the shaft ends, where the leakage is to the atmosphere, and between the stages, where the leakage flow can bypass the blade system and therefore reduce the efficiency. However, it is not possible to eliminate this leakage completely, because the shaft peripheral velocity is so high that any direct contact between it and a stationary member will generate enough heat to melt down the rubbing member and possibly badly distort the shaft as well. The seals therefore take the form of labyrinths with rather thin sharp teeth on at least one of the members (Fig. 4). In normal operation these members do not touch but run with a small clearance. The turbine is so designed that the running clearances in the shaft sealing labyrinths are as small as possible. In case of rubbing contact, which may happen accidentally, the sharp teeth can wear away without generating sufficient heat to distort the shaft. It is practical to construct the turbine with the radial clearance between the stationary and rotating teeth of the labyrinth no greater than .010–.015 in. on small and medium size turbines, and .030–.040 in. on large machines.

Shaft vibration and alignment. The shaft and bearings must be as free as possible of critical speeds in the range where the turbine is to operate. The shaft must also be stable and remain in balance once a good balance has been achieved. Part of the solution to this problem of balance lies in the design of the bearings and part also in the design of the turbine foundation and means for maintaining proper bearing alignment.

Governing. Turbines usually have two governors, one to control speed and one to limit possible overspeed. The speed control governor may be of any kind that will give a signal whose strength or position is a function of speed of the turbine. Flyball governors of refined design, and hydraulic governors whose essential part is a centrifugal pump, are both used extensively. In either case, for all

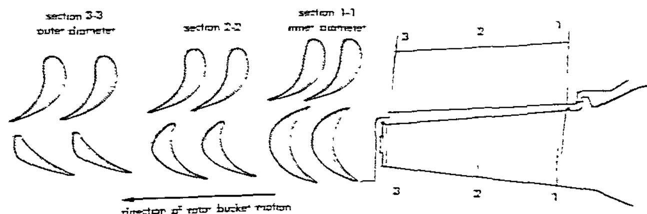


Fig. 5. Radial and tangential sections through the stationary and rotating rows of buckets of a stage from a large turbine.

but the smallest machines, the governor signal must be power relayed to furnish large enough forces to position the steam control valves accurately, closing them when the speed increases and opening them when the speed decreases. The speed-limiting governor, or overspeed governor, is so made as to go through its full stroke immediately if the turbine speed reaches a predetermined value above normal speed, usually about 10%. It is made to actuate a quick-closing emergency stop valve.

These two kinds of governors and their separate valve systems offer two lines of defense against a possible disastrous overspeeding should the turbine suddenly lose its load. Steam turbines cannot usually be designed to be safe at runaway speed. The governor systems are designed with great attention to their reliability.

The governors and the valves which they control are fast in action, the valves will usually travel from wide open to fully closed in a fraction of a second (*see* GOVERNOR).

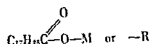
Oiling system. The turbine shaft runs at a high speed and its bearings must be positively and continuously supplied with oil; they will burn out quickly if the oil supply fails. The integrity of the oil system must be assured, two oil pumps, a main pump and a standby, driven by separate power sources, are usually provided on all but the smallest machines. The oil will also usually be used as a hydraulic power fluid to operate the steam control valves. It is customary to provide for closing the steam control valves to shut down the turbine if the oil supply pressure fails.

Aerodynamic design of blade system. The blading design for highest efficiency, especially for the larger sizes of turbines, draws upon modern aerodynamic theory. Classic forms of impulse and reaction blades merge together in the three-dimensional design required by modern concepts of loss-free fluid flow. To meet the theoretical steam flow requirements and to minimize centrifugal forces on the rotor blades and their attachments, the sections change in shape along a blade (Fig. 5).

Materials used in steam turbines. The materials used in the making of a steam turbine must be sound, strong and of the best quality. Casings or shells are almost always made of cast steel, and for the higher pressures and temperatures, it is alloyed with molybdenum, vanadium, chromium and sometimes other elements for greater strength. Small turbine casings for the lower range of temperatures may be of a good grade of cast iron, and the low-pressure casings of large machines may also be of this material. The rotors are steel forgings alloyed as may be necessary for greater strength with some of the materials mentioned above. The rotor may be built up, that is, made of separate wheels shrunk on a shaft, or it may be machined, wheels and all, out of one solid large forging, or it may be a number of disks welded together. The blades, both stationary and rotating, are usually of a low-carbon 12–13% chrome steel, which is very rust resistant. [P.H.K.]

Stearate

A salt (soap) or ester of stearic acid having the general formula



$$C_{17}H_{35}C(=O)—O—M \quad \text{or} \quad —R$$

and formed by replacing the carboxylic hydrogen by a metal (M) to give a salt, or by an organic radical (R) to give an ester. Stearates occur in nature chiefly as the glyceryl ester, found in substantial amounts in animal and vegetable fats. The esters of long chain alcohols are known as waxes. Other esters of monohydric and polyhydric alcohols are used in cosmetics, lacquers, and nonionic surface-active agents. Alkali-metal salts are water soluble, and with the similar oleates and palmitates are the major components of toilet and laundry soaps. Other metal salts are used in paints, waterproofing, pharmaceuticals, cosmetics, and fungicides. *See* FAT AND OIL, EDIBLE; FAT AND OIL, NONEDIBLE; SOAP AND DETERGENT; WAX, ANIMAL AND VEGETABLE. [E.H.H.]

Steel

Ferritic steels are those having a body-centered cubic crystal lattice at room temperature, as distinguished from austenitic steels which have a face-centered cubic lattice at room temperature. Ferritic steels constitute about 99% of the steel output of the United States. Steel is the basic industrial material of the twentieth century.

Known for its great strength, steel is an alloy of iron and carbon. Iron itself is not particularly strong, but its abundance and the remarkable increase in strength which becomes possible when

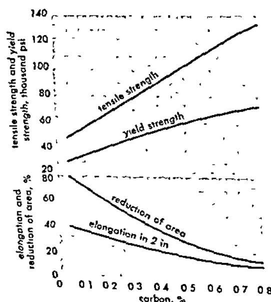


Fig. 1. Effect of carbon content on the tensile properties of hot-worked carbon steels.

iron is alloyed with carbon have given steel its preeminent place. About 92% of all steel is plain carbon steel; the remainder is alloy steel in which iron is alloyed not only with carbon but also with varying amounts of such alloying elements as manganese, nickel, chromium, molybdenum, and vanadium.

A large proportion of the plain carbon steels is used in the hot-rolled or forged condition, without any further treatment such as reheating and quenching. Figure 1 shows the effect of carbon content on the tensile properties of hot-worked carbon steels. The yield and tensile strengths increase with increasing carbon content, but the ductility, as indicated by the per cent elongation and reduction of area, decreases. Thus, to guard against brittle behavior the carbon content of steel must not be raised unduly. Steels containing about 0.25% carbon are classed as mild steels, those with 0.45% carbon as medium steels, and those with about 0.7% carbon and higher, as higher-carbon steels. The higher-carbon steels in particular, and some of the lower-carbon steels as well, are generally used in the heat-treated condition.

The tensile properties in Fig. 1 can be improved greatly by heat treatment, that is, by reheating the steel to above its critical temperature and cooling at various rates, particularly by rapid cooling as by water quenching.

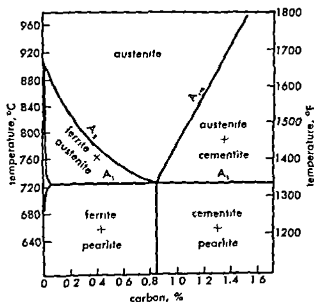


Fig. 2 Critical temperatures in plain carbon steels and constituents present in the various iron-carbon alloys upon slow cooling

Iron-carbon phases. Figure 2 shows the critical temperatures in plain carbon steels, and indicates the constituents present in the various iron-carbon alloys upon slow cooling. In slowly cooled steel, the carbon is present as a hard iron-carbon compound (iron carbide, Fe_3C) called cementite. In steels with less than about 0.8% carbon, the carbon is incorporated in the carbon-bearing constituent called pearlite, which consists of alternate lamellae of ferrite and cementite. In the steels with less than 0.8% carbon, the constituents present are ferrite and pearlite as indicated below the line A_1 in Fig. 2. In the steels containing more than 0.8% carbon, ferrite is no longer present; in these steels up to 0.8% of the carbon is present in the form of cementite lamellae in pearlite, while the rest of the carbon forms cementite boundaries about the pearlite grains.

The appearance of the ferrite, pearlite, and cementite constituents in a range of steels of increasing carbon content up to 1.2% carbon is shown in Fig. 3. As the content of carbon increases, the amount of pearlite increases, and above 0.8% carbon cementite appears. At the relatively low magnification of ×100 in Fig. 3, the alternate lamellae of ferrite and cementite of which pearlite is composed are not resolved, and the pearlite appears merely as dark areas. A magnification of ×500 and over is generally required to resolve pearlite clearly.

Figure 1 shows the strengths of slowly cooled carbon steels. A much higher order of hardness is obtainable if the steel is cooled very rapidly from above its critical temperature. Such rapid cooling, as by quenching the hot steel in cold water, is what is meant by hardening.

Figure 2 explains what takes place in the quench-hardening of steel. At temperatures above the line A_3 in Fig. 2 the steel is in the austenitic form, and not as ferrite and cementite. Hardening can occur mainly because the austenitic form of iron can hold a large amount of carbon in solution (2% carbon), whereas ferrite can hold only a minute amount of carbon in solution (0.035% carbon). If a steel is cooled slowly from above the critical temperature, the carbon can separate when the critical temperature is traversed, and the austenite changes to ferrite and cementite. However, when the steel is cooled very rapidly through the critical temperature, there is not time enough for the cementite to separate. The austenite transforms to ferrite but the cementite does not separate. The bulk of the cementite is, therefore, retained in the ferrite in supersaturated solution, although some of the cementite precipitates as an extremely fine (submicroscopic) dispersion. The extreme hardness of quench-hardened steel may be attributed to the supersaturated solution and fine dispersion of the cementite in the ferrite.

The constituent which forms in quench-hardened steel is called martensite. At about ×500 magnification, martensite can be distinguished from pearlite in slowly cooled steel by its acicular needlelike structure. Martensite is extremely hard and is the structure desired in steel for metal-cutting tools, files, and cutlery. Extreme hardness generally denotes brittleness. Thus, as-quenched martensite may be too brittle. Quench-hardened steel can be made tougher and less brittle by heating it after quenching. Such heating is called tempering. Within limits, the higher the tempering temperature, the softer, less brittle, and tougher is the steel. The quenched and tempered structure is, therefore, the commonly sought structure [illegible]t-treated

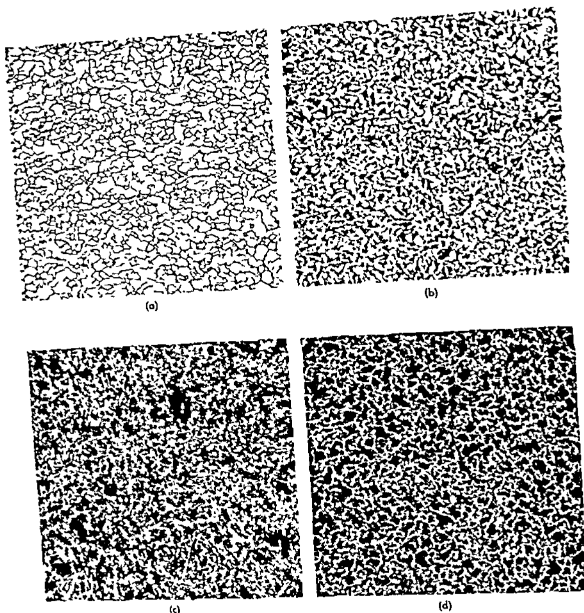


Fig 3 Micrographs X100 illustrating appearance of the constituents shown in Fig 2 (a) 0 06% carbon steel White is ferrite, grain boundaries appear as a dark network (b) 0 18% carbon steel White is ferrite, dark is pearlite. (c) 0.47% carbon steel. Same a (b), but larger amount of pearlite in this higher car bon steel (d) 1 2% carbon steel Dark is pearlite white grain boundaries are cementite

steel The tempering temperature is determined by the hardness required in the finished article For extreme hardness the tempering temperature may be about 400°F for great toughness the tempering temperature may be as high as 1200°F. During tempering the cementite separates from the martensite and at the higher tempering temperatures it coalesces into globular particles large enough to be resolved in the microscope

Just as the carbon content of the steel determines the strength in the slowly cooled steels, as shown in Fig 1 so also does the carbon content determine the hardness and strength in the quench-hardened steels. This is shown in Fig 4, the Rockwell C penetration hardness of the steels being indicated at the left and the corresponding tensile strength at the right The quenched hardness rises rapidly with carbon content A steel low in carbon as 0 20% has a Rockwell C hardne of 50, corresponding to about 245,000 psi tens strength At 0 60% carbon the hardness is 65 Rockwell C, corresponding to about 450,000 psi tens strength There is only a relatively small furth increase in hardness with further increases in c bon content

Alloy steels. In general the alloy elements alloy steel make only a minor contribution to attainable hardness after quenching; the hardn depends mainly on the carbon content. The portant effect of the alloying elements lies no increasing the hardness but in slowing down rate at which austenite transforms to ferrite

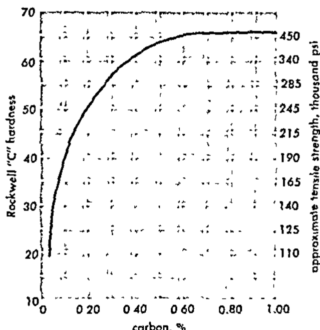


Fig. 4. Relation of hardness and tensile strength to carbon content in steels rapidly quenched from above the critical temperature.

cementite during cooling through the critical temperature region. As a result, instead of requiring extremely fast cooling to produce martensite as in plain carbon steel, very much slower cooling is sufficient to produce martensite in alloy steel. This characteristic of the alloying elements in alloy steel determines the hardenability of the steel.

A 1-in diameter bar of plain carbon steel, heated to above the critical temperature and then quenched in cold water, will cool rapidly enough to form martensite only at the surface. Martensite will be formed only to a depth of about ⅛ in In the interior of the plain carbon steel bar the rate of cooling even in cold water will be too slow to form martensite On the other hand, in a bar 1 in in diameter of alloy steel similarly water-quenched or even more slowly cooled as by quenching in oil, martensite will form throughout the bar It is not always desirable to quench steel parts drastically in water, because this may cause warping and cracking; milder quenching, as with oil, may therefore be preferred.

The main function of the alloying elements is to increase the hardenability, that is, to make possible deep or full martensitic hardening throughout large sections. Steel that is fully martensitic after hardening is definitely tougher after tempering to a given hardness than steel that is only partly martensitic after hardening and then tempered to the same given hardness.

Before 1940, knowledge about hardenability was obtained by quenching round steel bars and observing the depth of hardening, or, preferably, by making bars of varying diameters and testing to see which would harden clear through. This was a cumbersome task and the concept of hardenability remained essentially qualitative. In more recent years, however, the Jominy test procedure has come into extensive use. This consists of quenching the end of a 1-in round bar 4 in long and determining the hardness at intervals along the length of the bar, beginning with the quenched end The distance from the quenched end to which the steel hardens is thus obtained. These results can be related to the cooling rates at the various distances from the quenched end and this in turn to the diameter of the bar which will fully harden to its center. In this way fairly precise quantitative multiplying factors, indicating the degree by which a given per cent of an alloying element will increase the hardenability, have been determined Such multiplying factors are given below for the following alloying elements, arranged in the order of their increasing effect on hardenability, when 0.5% of the alloying element is present in a steel of the same carbon content and grain size: Ni, 1.18; Si, 1.35, Cr, 2.08; Mo, 2.50; and Mn, 2.67.

Such precise methods of determining hardenability have been of aid in demonstrating the remarkably strong effect of boron In medium carbon steels, 0.001% boron increases the hardenability as much as 0.25% of chromium The explanation for

Analyses of some standard engineering steels*

Designation, SAE number	C%	Mn%	P%	S%	Si%	Ni%	Cr%	Mo%
C 1006	0.08 max	0.25-0.40	0.040 max	0.050 max				
C 1012	0.10-0.15	0.30-0.50	0.040 max	0.050 max				
B 1111	0.08-0.13	0.70-1.00	0.07-0.12	0.10-0.15				
C 1111	0.08-0.13	0.60-0.90	0.045 max	0.16-0.23				
A 2317	0.15-0.20	0.40-0.60	0.040 max	0.040 max	0.20-0.35	3.25-3.75		
A 3140	0.38-0.43	0.70-0.90	0.040 max	0.040 max	0.20-0.35	1.10-1.40	0.55-0.75	
A 4068	0.64-0.72	0.75-1.00	0.040 max	0.040 max	0.20-0.35			0.20-0.30
A 5120	0.17-0.22	0.70-0.90	0.040 max	0.040 max	0.20-0.35		0.70-0.90	
E 52101	0.95-1.10	0.25-0.45	0.025 max	0.025 max	0.20-0.35		1.30-1.60	
A 8612	0.10-0.15	0.70-0.90	0.040 max	0.040 max	0.20-0.35	0.40-0.70	0.40-0.60	0.15-0.25
E 9320	0.18-0.23	0.45-0.65	0.025 max	0.025 max	0.20-0.35	3.00-3.50	1.00-1.40	0.08-0.15
NE 9768	0.64-0.72	0.50-0.80	0.040 max	0.040 max	0.20-0.35	0.40-0.70	0.10-0.25	0.15-0.25
A 9262	0.55-0.65	0.70-1.00	0.040 max	0.040 max	1.80-2.20		0.25-0.40	

* Charles M. Parker, Standard engineering steels, *Metals and Alloys*, September, 1945

this exceptionally strong effect of boron appears to be that boron concentrates in the grain boundaries of the steel, a particularly favorable location for slowing the rate of transformation of austenite to ferrite and pearlite during quenching. The analyses of some commonly used steels are shown in the table.

The foregoing briefly reviews some of the main generalizations of the physical metallurgy of the ferritic steels. Not only such general principles, but also the technical skills and inventiveness of the steel producing and -consuming industries have played their part in the development of the myriad of different steel products. Each of these products has its special qualities, and for expertness different men have to specialize in each of them. Thus, some metallurgical engineers specialize in sheet steel products, others in structural steel, in plate, in pipe, in the various kinds of carbon and alloy steel bars, in tool and die steel, in wire and wire products, in railroad rails, wheels, and axles, and in large forgings. See FERROALLOY; HEAT-TREATMENT (METALS AND ALLOYS); IRON ALLOYS; METAL, MECHANICAL PROPERTIES OF. [S.F.P.]

Bibliography: American Society for Metals, *Metals Handbook*, 1948; D. K. Bullens, *Steel and Its Heat Treatment*, 5th ed., vols. 1–3, 1948–1949. S. Epstein, *The Alloys of Iron and Carbon*, vol. 1, 1936; M. E. Shank, *Control of Steel Construction to Avoid Brittle Failure*, 1957; F. T. Sisco, *The Alloys of Iron and Carbon*, vol. 2, 1937; U.S. Steel Corporation, *The Making, Shaping, and Treating of Steel*, 7th ed., 1957.

Steel manufacture

The manufacture of steel is a tremendous industry with extensive ramifications resulting from the complex interrelation of technical and economic considerations. The world production for 1957 was 322,000,000 tons. The United States has been the largest single producer since 1890 and produced 112,[illegible]

ag [illegible]

the [illegible] term generally refers to the product of the iron blast furnace. See IRON (EXTRACTION FROM ORE). This is a complex alloy containing about 6% of other common elements whose distribution depends on the raw materials and operation of the particular blast furnace, but which can be represented by a typical analysis: carbon (C), 4.5%, silicon (Si), 1.5%, manganese (Mn), 0.8%, phosphorus (P), 1.0%, sulfur (S), 0.03%, and many minor impurities. The variation of the Si and P in particular makes this product more or less suitable for a specific steelmaking process. As this molten product comes from the blast furnace at a temperature of about 1400°C (2600°F), it is called hot metal and is sent directly to the steelmaking operations to be described. If the product is solidified in small molds for convenient handling, it is called pig iron, or solidified in useful shapes such as car wheels, it is called cast iron. Cast iron is relatively brittle and has only limited commercial applications accounting for less than 10% of all the iron made. Thus, the steelmaking processes consume most of the iron made, and nearly all of this is in the form of hot metal. Companies which use their blast-furnace product in this way are referred to as integrated companies. However, this blast-furnace iron provides only about one-half of the metallic raw material for the steel industry. The other half is obtained from scrap produced in subsequent manufacturing operations and from obsolescence of steel products. Because the iron units in this scrap are cheaper than from the blast furnace, the ability of a steelmaking process to use more or less scrap advantageously is another important characteristic of the process.

The steels designated as plain carbon steels generally contain less than 0.8% C, minimum amounts of P, S, and suitably adjusted amounts of Si and [illegible]

[illegible] than 5%, are usually made by the same processes as plain carbon steels. The high-alloy steels, such as stainless containing 18% Cr and 8% Ni, require special processes.

A special group of materials known as ferroalloys constitute the major source of alloy additions for all grades of steel. However, in many cases, these alloying metals are also available and used in relatively pure forms, such as electrolytic nickel. See FERROALLOY.

The melting point of pure iron is 1535°C. This is modified somewhat by the alloys present in commercial steels, but in order to provide proper pouring conditions for both ingots and castings, it is generally necessary to reach temperatures in the range 1550–1650°C in all steelmaking processes. These high temperatures place serious limitations on the refractories which can be used. They also require the most efficient use of heat, whether it is derived from combustion of fuel, electricity, or thermochemical reaction. Iron and steel products have been used by man since antiquity, but steels could not be made in tonnage quantities before 1850.

Thus, the various modern steelmaking processes have been developed because of their combined attributes for the use of available raw materials, the [illegible]

e [illegible]

T [illegible]

t [illegible] processes.

Classification of processes. From a chemical standpoint, all steelmaking processes may be classified as acid or basic, depending upon the refractory and slag combination, each having particular attributes with regard to the refining which it can accomplish.

Acid processes use silica (SiO_2) refractories throughout, and are able to accommodate slags which become saturated with this component under

operating conditions. These acid systems can be used to eliminate carbon, manganese, and silicon from the charge, but require select raw materials within final steel specifications for phosphorus and sulfur.

Basic processes use magnesite (MgO) refractories in the portions of the furnace which contact molten slag and metal, and accommodate lower silica slags with compensating amounts of lime (CaO). These systems can eliminate carbon, manganese, and silicon as effectively as the acid systems and will also eliminate phosphorus and appreciable amounts of sulfur. This gives basic systems a decided advantage in flexibility with regard to raw materials consumed and grades of steel produced

With regard to engineering features there are three main types of processes: (1) pneumatic, in which all heat is derived from the initial heat content of the charge materials, principally molten, and the thermochemical balance of the refining reactions; the selective oxidation of the refining is accomplished by blowing air or commercial oxygen; (2) open hearth, in which the major source of heat is the combustion of fuel (usually gas or oil); this in turn depends for its success on the regenerative principle of preheating air in order to attain steelmaking temperatures efficiently; and (3) electric, in which the major source of heat is electric current (arc or resistance or both); since this heat can be produced in the presence or absence of oxygen, electric furnaces can operate in neutral or nonoxidizing atmospheres or vacuum, and thus are either preferred or required for alloys with significant amounts of easily oxidized elements.

In order to use available raw materials and heat sources effectively for particular grades of steel, steelmaking processes of the pneumatic, open-hearth, and electric types exist, and any of these may be either acid or basic in their chemistry. In many cases, there is an overlapping or combination of the above principles in a single process, but the indicated classification is followed in describing the process and its product.

Steel production in the United States

Process	Production 1957, net tons	Capacity 1958, net tons
Basic open hearth	101,027,725	121,502,400
Acid open hearth	630,051	819,430
Bessemer	2,475,138	4,027,000
Electric and crucible*	8,582,082	13,312,740
Oxygen		1,081,000
Total	112,714,996	140,742,570

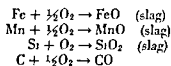
* For 1957, steel made by the oxygen process was included as crucible steel

The relative importance of the main commercial processes according to tonnage is given in the table. The principal characteristics of the individual processes will be described.

PNEUMATIC PROCESSES

In the United States, there is a limited amount of low-P ore available. This type of ore produces a blast-furnace product suitable for the Bessemer process, based on the original practice developed in England by Sir Henry Bessemer in the years following his original patent in 1857. The same process was proved in the United States by William Kelly after his original idea in 1847. This is an acid process, and was historically the first high-tonnage process for steel production. It will be described first because, in addition to this historical position, it is the simplest major process. From the beginning, this process was able to produce steel from hot metal in 10- to 25-ton heats at an average rate of nearly 1 ton/min. Although no scrap is required, it normally uses 12–15% scrap. This low scrap requirement is an advantage when the supply is limited and was a major reason for the initial dominance of this process. Suitable hot metal for the process should analyze C, 4.00–4.50%; Si, 1.10–1.50%; Mn, 0.40–0.70%; P, 0.09% maximum; and S, 0.03% maximum.

Converters. The typical vessel or converter, shown in Fig. 1, is a refractory-lined steel shell with tuyeres in the bottom and open at the top. The vessel is mounted on trunnions and is provided with mechanical means for tilting. The air for the process is blown through one hollow trunnion and is distributed to the tuyeres through the wind box at a pressure of about 30 psi. In a typical blow, the hot metal is charged while the vessel is tilted to the horizontal position to keep the tuyeres clear. The blast is turned on as the vessel is righted, and the air bubbling through the melt, about 18–24 in deep, provides the oxygen required for the refining reactions



$$Fe + \tfrac{1}{2}O_2 \rightarrow FeO \quad \text{(slag)}$$
$$Mn + \tfrac{1}{2}O_2 \rightarrow MnO \quad \text{(slag)}$$
$$Si + O_2 \rightarrow SiO_2 \quad \textit{(slag)}$$
$$C + \tfrac{1}{2}O_2 \rightarrow CO$$

These occur approximately in the order listed, although there is considerable overlapping. These reactions are exothermic and provide enough heat to raise the temperature from that of the charge (about 1350°C) to that of the product (about 1600°C). This temperature rise can be controlled *by regulation of the hot-metal analysis (Si content* being the most important) and the amount of scrap added during the blow. The flame emitted from the mouth of the vessel changes its color and luminosity in a manner which allows the operator (blower) to judge the progress of the refining reactions, to stop the operation at a suitable endpoint, and to pour the heat into a transfer ladle by tilting the vessel.

The oxides formed in the first three reactions combine to form a silica-saturated slag. This will analyze approximately 50% SiO_2 and 50% (MnO + FeO). The silica requirement in excess of that provided by the Si from the hot metal is

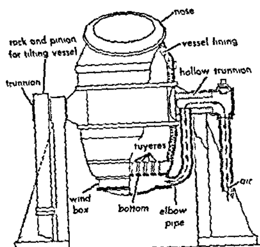


Fig 1 Schematic cutaway view of Bessemer converter (From United States Steel Corp, The Making, Shaping, and Treating of Steel, 7th ed., 1957)

necessarily obtained from the refractory lining of the vessel The entire blowing time is only about 15 min, so a great deal of the success of the process depends upon the blower's skill The time is too short to allow for control by sampling and analysis.

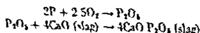
Applications. Although many variations are possible, the process is typically limited to plain carbon steels which contain less than 0.30% C, and which, because of the intimate contact with air, will be higher in nitrogen content than the open-hearth product These, and associated features, make it most suitable for applications which do not require maximum ductility and in which work hardening is desirable The steel welds well, and special grades, such as high-sulfur screw stock, have exceptional machinability A significant fraction of the hot metal refined by this process is passed on to the basic open hearth for further refining

Additional operations of deoxidation, alloy additions and teeming (pouring from a transfer ladle into ingot molds) are required to complete the process. These are essentially the same as will be described in connection with the more important basic open-hearth process.

The basic Bessemer or Thomas process is important in Europe and was developed to use the iron from high phosphorus ores which cannot be refined by acid slags The refractory lining is made of magnesite (MgO), and an afterblow eliminates the phosphorus by the reactions

$$2P + 2.5O_2 \rightarrow P_2O_5$$
$$P_2O_5 + 4CaO\ (slag) \rightarrow 4CaO \cdot P_2O_5\ (slag)$$

The lime, CaO required for this purpose is added with the charge and is dissolved in the slag throughout the blow so that the final slag has the necessary CaO SiO_2 ratio of about 3 The important differences between the acid and basic processes are with regard to the raw material requirements rather than the type of product produced.



BASIC OPEN HEARTH PROCESS

The basic open hearth has been the dominant tonnage steel producer in the United States since the beginning of the twentieth century, and it promises to remain in this position for the foreseeable future, even though it may lose some ground to modern variations of converter and electric-furnace processes. Historically, the conventional reverberatory-type furnace was unable to generate sufficiently high temperatures for proper refining of steels, and only the limited (500 lb or less per day) production of the special grade known as wrought iron was possible. In 1858, Karl Siemens built an ex- [illegible]

sential design of modern open-hearth furnaces. [illegible]

within a few years of each other. The initial rapid development of the Bessemer (pneumatic) process was a result of its inherent simplicity. The final dominance of the open hearth is the result of its ability to use scrap advantageously, its greater flexibility in use of other raw materials, and its ability to make a wider range of steels. It can be operated as either an acid or basic process by selection of refractories, and in either case, a wide variation in practices is possible.

Furnaces. The features of furnace construction and the regenerative principle are shown in Fig. 2. The hearth proper is lined with magnesite refractory for a basic furnace or silica for an acid furnace. The roof is normally made of silica brick with sprung arch construction, but there is a trend toward basic brick with suspended-arch construction. The selection of roof refractories is independent of the hearth construction, because this does not contact molten slag There is great variation in the refractories used in the remainder of the system, but this will not be considered here as it is not directly related to the steelmaking process. Burners are provided at both ends of the furnace and are operated alternately on a controlled cycle, usually 10–15 min The usual fuels are natural gas or oil, but this selection depends on availability and economy and is not otherwise limited. In order to attain useful temperatures in the furnace (above 1650°C) with minimum fuel cost, the air for combustion is preheated by the regenerative system. For example, when fuel is burned at one end, the dampers are adjusted so the air for combustion first passes through the hot checkers at this end. The checkers are an open latticework of refractory brick The flame and products of combustion pass over the hearth area, through the checker system on the opposite end and out the stack. The dampers are then reversed and fuel burned at the opposite end Thus, by selection of a suitable cycle, high temperatures are achieved with minimum fuel con-

operating conditions. These acid systems can be used to eliminate carbon, manganese, and silicon from the charge, but require select raw materials within final steel specifications for phosphorus and sulfur.

Basic processes use magnesite (MgO) refractories in the portions of the furnace which contact molten slag and metal, and accommodate lower silica slags with compensating amounts of lime (CaO). These systems can eliminate carbon, manganese, and silicon as effectively as the acid systems and will also eliminate phosphorus and *appreciable amounts of sulfur This gives basic* systems a decided advantage in flexibility with regard to raw materials consumed and grades of steel produced.

With regard to engineering features there are three main types of processes: (1) pneumatic, in which all heat is derived from the initial heat content of the charge materials, principally molten, and the thermochemical balance of the refining reactions; the selective oxidation of the refining is accomplished by blowing air or commercial oxygen; (2) open hearth, in which the major source of heat is the combustion of fuel (usually gas or oil); this in turn depends for its success on the regenerative principle of preheating air in order to attain steelmaking temperatures efficiently; and *(3) electric, in which the major source of heat is* electric current (arc or resistance or both); since this heat can be produced in the presence or absence of oxygen, electric furnaces can operate in neutral or nonoxidizing atmospheres or vacuum, and thus are either preferred or required for alloys with significant amounts of easily oxidized elements.

In order to use available raw materials and heat sources effectively for particular grades of steel, steelmaking processes of the pneumatic, open-hearth, and electric types exist, and any of these may be either acid or basic in their chemistry In many cases, there is an overlapping or combination of the above principles in a single process, but the indicated classification is followed in describing the process and its product

Steel production in the United States

Process	Production 1957, net tons	Capacity 1958, net tons
Basic open hearth	101,027,725	121,502,400
Acid open hearth	630,051	819,430
Bessemer	2,475,138	4,027,000
Electric and crucible*	8,582,082	13,312,740
Oxygen		1,081,000
Total	112,714,996	140,742,570

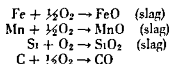
* For 1957, steel made by the oxygen process was included as crucible steel

The relative importance of the main commercial processes according to tonnage is given in the table. The principal characteristics of the individual processes will be described

PNEUMATIC PROCESSES

In the United States, there is a limited amount of low-P ore available. This type of ore produces a blast-furnace product suitable for the Bessemer process, based on the original practice developed in England by Sir Henry Bessemer in the years following his original patent in 1857. The same process was proved in the United States by William Kelly after his original idea in 1847. This is an acid process, and was historically the first high-tonnage process for steel production. It will be *described first because, in addition to this historical* position, it is the simplest major process. From the beginning, this process was able to produce steel from hot metal in 10- to 25-ton heats at an average rate of nearly 1 ton/min. Although no scrap is required, it normally uses 12–15% scrap This low scrap requirement is an advantage when the supply is limited and was a major reason for the initial dominance of this process. Suitable hot metal for the process should analyze C, 4 00–4 50%; Si, 1.10–1.50%; Mn, 0 40–0.70%; P, 0.09% maximum; and S, 0 03% maximum

Converters. The typical vessel or converter, shown in Fig 1, is a refractory-lined steel shell with tuyeres in the bottom and open at the top. The vessel is mounted on trunnions and is provided with mechanical means *for tilting The air for the process* is blown through one hollow trunnion and is distributed to the tuyeres through the wind box at a pressure of about 30 psi In a typical blow, the hot metal is charged while the vessel is tilted to the horizontal position to keep the tuyeres clear. The blast is turned on as the vessel is righted, and the air bubbling through the melt, about 18–24 in. deep, provides the oxygen required for the refining reactions



$$Fe + \tfrac{1}{2}O_2 \rightarrow FeO \quad \text{(slag)}$$
$$Mn + \tfrac{1}{2}O_2 \rightarrow MnO \quad \text{(slag)}$$
$$Si + O_2 \rightarrow SiO_2 \quad \text{(slag)}$$
$$C + \tfrac{1}{2}O_2 \rightarrow CO$$

These occur approximately in the order listed, although there is considerable overlapping These reactions are exothermic and provide enough heat to raise the temperature from that of the charge (about 1350°C) to that of the product (about 1600°C). This temperature rise can be controlled by regulation of the hot-metal analysis (Si content being the most important) and the amount of scrap added during the blow The flame emitted from the mouth of the vessel changes its color and luminosity in a manner which allows the operator (blower) to judge the progress of the refining reactions to stop the operation at a suitable endpoint, and to pour the heat into a transfer ladle by tilting the vessel

The oxides formed in the first three reactions combine to form a silica-saturated slag This will analyze approximately 50% SiO_2 and 50% (MnO + FeO). The silica requirement in excess *of that provided by the Si from the hot metal is*

Fig. 1 Schematic cutaway view of Bessemer converter. (From United States Steel Corp., *The Making, Shaping, and Treating of Steel, 7th ed.*, 1957)

necessarily obtained from the refractory lining of the vessel. The entire blowing time is only about 15 min so a *great deal of the success of the* process depends upon the blower's skill. The time is too short to allow for control by sampling and analysis.

Applications. Although many variations are possible, the process *is* typically limited to plain carbon steels which contain less *than 0.30% C*, and which, because of the intimate contact with air, will be higher in *nitrogen content* than the open-hearth product. These, and associated features, make it most suitable for applications *which do not* require maximum ductility and in which work *hardening* is desirable. The steel welds well, and special *grades*, such as high-sulfur screw stock, have *exceptional* machinability. A *significant* fraction of the hot metal refined by this process is passed *on* to the basic open hearth for further refining.

Additional operations of deoxidation, alloy additions and teeming (pouring from a transfer ladle into ingot *molds*) are required to complete the process. These are essentially the same as will be described in connection with the more important basic open hearth process.

The basic Bessemer or Thomas process is important *in* Europe and was developed to use the iron from high phosphorus ores which cannot be refined by acid slags. The refractory *lining* is made of magnesite (MgO), and an afterblow eliminates the phosphorus by the reactions

$$2P + 2.5O_2 \rightarrow P_2O_5$$
$$P_2O_5 + 4CaO\ (slag) \rightarrow 4CaO\ P_2O_5\ (slag)$$

The lime, CaO, required for this purpose is added with the charge and is dissolved in the slag throughout the blow so that the *final* slag has the necessary $CaO\ SiO_2$ *ratio* of about 3. The *important* differences between the acid and basic processes *are* with regard to *the raw* material *requirements* rather than the type of product *produced*.

BASIC OPEN HEARTH PROCESS

The basic open hearth has been the dominant tonnage steel producer in the United States since the beginning of the twentieth century, and it promises to remain in this position for the foreseeable future, even though it may lose some ground to modern variations of converter and electric-furnace processes. Historically, the conventional reverbera*tory-type furnace* was *unable to generate sufficiently* high temperatures for proper refining of steels, and only the limited (500 lb or less per day) production of the special grade known as wrought iron was possible. In 1858, Karl Siemens built an experimental furnace making use of his regenerative principle, and together with his brother, Frederick, developed it within the next few years into the essential design of *modern open-hearth furnaces*. Thus, both the pneumatic and the open-hearth processes, which form the basis of the modern era of low-cost, high-tonnage steel, were established within a few years of each other. The initial rapid development of the Bessemer (pneumatic) process was a result of its inherent simplicity. The *final* dominance of the open hearth is *the result of its* ability to use scrap advantageously, its greater flexibility in use of other raw materials, and its ability to make *a* wider range of steels. It can be operated as *either an acid* or basic process by selection of refractories, and in *either case, a* wide variation in *practices is possible.*

Furnaces. The *features of* furnace construction and the regenerative principle are shown in Fig. 2. *The hearth proper* is lined with magnesite refractory for a basic furnace or silica for an acid *furnace*. *The roof* is normally made of silica brick with sprung arch [illegible] wa [illegible] Th [illegible] of [illegible] does not contact molten slag. There is great variation in the refractories *used in the* remainder of the system, but this will not be considered here as it is *not* directly related to the steelmaking process. Burners are provided at both ends *of the* furnace and are *operated alternately* on a controlled cycle, usually 10–15 min. The *usual* fuels are natural gas or oil [illegible] ecc [illegible] att [illegible] surface (above 1650 C) with minimum *fuel* cost, the air for combustion is *preheated by the* regenerative system. For example, *when* fuel is burned at *one* end, the dampers are adjusted so the air for combustion first passes through the hot checkers at this end. The checkers are *an open* latticework of refractory brick. The flame and products of combustion pass over the hearth area, through the checker system *on the* opposite end and out *the stack*. The dampers *are* then reversed and fuel burned at the opposite end. Thus, by selection of a suitable cycle, high temperatures are achieved with minimum *fuel* con-

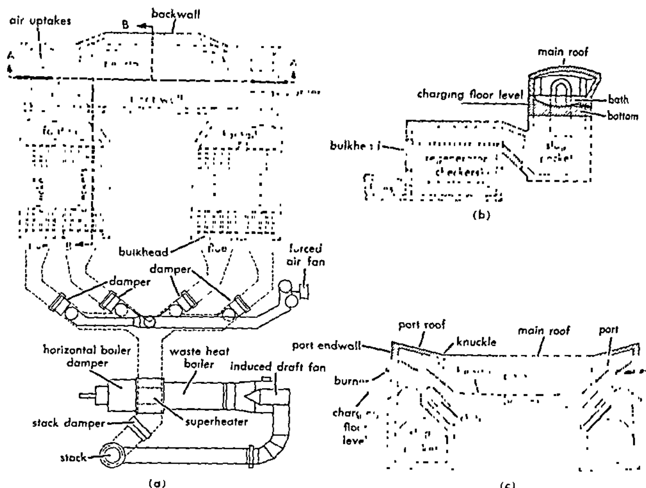


Fig. 2. Schematic sections of liquid-fuel-fired open-hearth furnace. (a) Sectioned schematic plan. (b) Vertical section across B-B. (c) Vertical section across A-A *(From United States Steel Corp., The Making, Shaping, and Treating of Steel, 7th ed., 1957)*

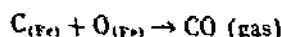
sumption. Typical modern furnaces have a capacity of 200–300 tons per heat.

The making of a typical heat of plain C steel by the basic open-hearth process will be outlined. The furnaces are operated continuously in order to conserve heat, minimize spalling of refractories, and produce maximum tonnage. After the previous heat has been drained from the hearth, grain magnesite is blown or shoveled into the areas where erosion has occurred, and then the materials for the next heat can be charged. The solids are previously weighed into charging boxes and brought into the furnace area on charging buggies by the mill locomotive. These boxes can be lifted, thrust through the furnace door, and dumped by the charging machine. The mechanization of these operations is an [illegible]

[illegible] charge mate- [illegible] Fe_2O_3), and scrap steel, and they are piled into the furnace in this order. The scrap constitutes about one-half the total iron units required for the heat. It will be a mix of light, voluminous and heavy, dense material and is distributed to achieve rapid heat absorption, economy of space, and so as to hold the lime and ore on the bottom as long as possible. The amounts of lime and ore are calculated to give the correct analysis of metal and slag at the end of the heat and will thus vary according to the type of scrap available and the grade of steel being made. As this charge is heated and melted, additional iron oxide is formed, because the furnace atmosphere must contain an excess of oxygen in order to maintain efficient combustion. Thus, as the scrap melts, it will be refined by the same oxidizing reactions already indicated for the pneumatic processes, and the slag formed will be highly oxidizing. When this scrap is partially melted, the remaining units of iron are added as hot metal by pouring from a transfer ladle through a trough placed in one of the charging doors.

An understanding of the remainder of the process requires a more detailed consideration of the chemistry involved. In order to make this applicable to all grades of steel and all processes, some important relations will be generalized. These can be made quite quantitative for any particular situation, but it should be recognized that this is an intricate and complicated undertaking because steelmaking involves heterogeneous reactions (gas-metal, gas-slag, and slag-metal) between phases which are not necessarily in equilibrium with each other with regard to either temperature or chemistry. The various stages and types of steelmaking involve the selective control of oxidizing reactions. The usual sources of oxygen are from the air or products of com-

bustion (CO_2 *is oxidizing toward Fe*); slag which contains large amounts of iron oxide usually represented as FeO; or ore represented as Fe_2O_3. Whatever the source, there will be an effective oxygen pressure and the molten bath of metal will dissolve



corre- This other elements in the metal such as $C_{(Fe)}$ as represented by the reaction

$$C_{(Fe)} + O_{(Fe)} \rightarrow CO \text{ (gas)}$$

Similar reactions occur with the Si, Mn, and other elements in the molten bath, as already shown under the Bessemer process. The equilibria for all of these reactions are of course interrelated; however, for present purposes, it is useful to consider the CO *reaction* in the bath as the dominant one during the refining period of the steelmaking process. The relations between the carbon and oxygen dissolved in the *bath are shown schematically* in Fig. 3, which indicates that for a given carbon content the amount of oxygen *in* the bath will increase with temperature, whereas for a given temperature the amounts of carbon and oxygen *in* the bath are inversely related.

Slag and steel. The above information may be applied to a heat under *way in the basic* open hearth furnace. *Pure iron* melts at 1535°C, whereas the hot metal melts at about 1130°C and *is probably* no hotter than 1300–1350°C. Thus the low-C *steel* scrap has a higher melting point than the hot *metal* (high C) when *most of the* scrap has been melted, so that this metal bath and the slag which *has formed will be high in* oxygen in relation to the *hot metal* being added. *This* results in the rapid refining *of the* Si, Mn, P, and C *in the system*. *In* particular the bubbling of CO causes the slag to become *foamy and voluminous*, and a large amount of it is flushed out of the furnace through either a slag notch at the back or an *open* door in the front, or both. *Thus this flush* slag *removes* from the system important *amounts* of the Si, Mn and P oxides which result *from* the refining reactions, *thereby* reducing the weight of slag *remaining* in the furnace and facilitating the heat transfer and slag control required in the *rest* of the process. After melting of the scrap starts, a balance must be maintained between the *temperature* increase of the bath and its C content so that the bath remains molten at all times. The stirring of the bath by CO *evolution* is important in achieving the necessary heat transfer, which is *often* the limiting rate of *the* process. The reactions during this flush period may *become* dangerously violent, and their control by proper timing of the hot-metal addition *requires* great skill by *the melter*.

During the *meltdown* period the limestone on the hearth is calcined

$$CaCO_3 \rightarrow CaO + CO_2$$

and as the last heavy scrap melts, this stone floats up into the slag, where it *must* be dissolved so *that* the final slag will have some predetermined CaO content. *Depending* again on the quality of the *starting materials* and the *specifications for* the steel being made, this exact composition will vary, but *it* will usually result in a final ratio of $CaO/SiO_2 = 2.5–3$. This ratio is a simple measure of slag basicity. This high-lime content is required to achieve the dephosphorization which is characteristic of all basic steelmaking. *It also* permits the *essential elimination* of significant amounts of sul*fur which* are introduced by the hot metal and also by the fuel. The equilibria are such that when the required slag basicity and C content of the metal bath are reached, the heat is ready to take out of the furnace through the tap hole at the back where a refractory-lined runner directs it into a transfer ladle. As the metal runs into the ladle, additions are made to adjust the final chemistry of the heat with regard to carbon and other elements, as well as oxygen control, according to the type of steel being made. Compared to the Bessemer, the open hearth is a slow process with a total charge-to-tap time of about 10 hours. The rate of refining at the end of the heat can be well controlled in order to allow for adequate sampling and analysis of the bath prior to tapping, thus permitting close adjustment to any carbon or other specification, providing the final solid additions are small enough to avoid chilling the heat in the ladle. In general, additions of up to 10% *alloy content can be made*.

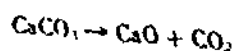
OXYGEN CONTROL

The final control of the oxygen content is critical, and this section can be applied, with only minor modification to the finishing of any heat of steel made by any process. Figure 3 shows that, as the *steel with* a given carbon specification cools, the $O_{(Fe)}$ which was required to achieve this carbon is *greater than that* in equilibrium with the same carbon at a lower temperature. Unless *remedial* measures are taken, this lowering of $O_{(Fe)}$ content will occur by two processes:

1. Separation of droplets of an oxide high in *iron whose exact* composition will depend on the *other elements in* the steel. Depending upon the conditions of *solidification*, these droplets may separate as a slag or be trapped during *freezing*, *in* which case they are called inclusions. These inclusions are more or less harmful, according to their composition and size and the use to be made of the steel.

2. *The evolution* of CO bubbles by the adjustment of the $C_{(Fe)}$-$O_{(Fe)}$ reaction to its temperature equilibrium. Particularly *in* low-C steels ($C <$ 0.20%), this would *generally be* so violent as to cause the steel to boil out of the molds during solidification, to cause *damage* to facilities, and to produce an unsound, worthless product.

According *to the final* type of oxygen control, steels are classified as fully deoxidized or "killed", "rimmed" or "open," which involves the minimum practical amount of deoxidation, and "semikilled,"



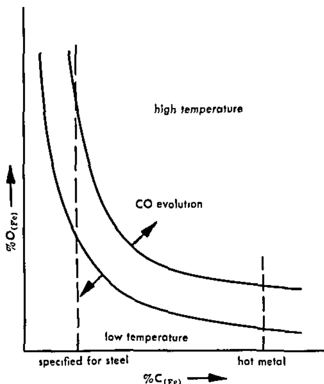


Fig 3. Generalized relations between carbon and oxygen dissolved in molten iron

which covers a range of conditions between killed and rimmed steels.

Killed steel. This steel is selected for sound internal structure. It is made by adding in the ladle some other element such as aluminum whose equilibrium with oxygen is such that the $O_{(Fe)}$ is lowered to a value which prevents the formation of CO. Other common deoxidizers, used separately or in combination, usually as ferroalloys, are Mn, Si, and Ti. The killed steel is then teemed into a big-end-up mold, as indicated in Fig 4 The molds are made of cast iron and provided with a hot top (an insulated upper section) which is designed to ensure that this will be the last region to freeze. The volume of the molten steel at the freezing point is about 8% greater than that of the solid The loss of heat during solidification is through the mold wall and by radiation from the open top. Thus, if the ingot mold had straight sides and no hot top, a rigid outer skin would freeze early in the process, and the difference in volume between liquid and solid would appear as a shrinkage cavity or pipe along the center line of the ingot. The tapered mold with the big end up and hot top is designed to ensure progressive freezing from bottom to top so that liquid steel can continue to feed into the center shrinkage zone as long as possible. A small pipe in the hot top area is inevitable, but this can be cropped and most of the ingot used as sound metal. Thus, killed steels eliminate the problem of CO evolution during freezing and minimize the objectionable features of the resultant pipe by suitable mold design. The treatment required to accomplish this creates deoxidation products which determine the final steel quality

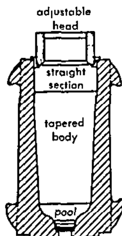


Fig 4. Typical big-end-up mold with hot top, used in killed steel. (From J. L. Bray, *Ferrous Process Metallurgy*, Wiley, 1954)

All of the elements used as deoxidizers influence the solubility of oxygen in the molten steel in a manner similar to that indicated for C in Fig. 3 In order to be effective, they must lower the $O_{(Fe)}$ below that of the specified carbon. It is apparent that if a solid deoxidation product such as aluminum oxide, Al_2O_3, is formed when the deoxidizer is added to the ladle, it may have a chance to float out of the metal before it freezes. However, the equilibrium is continually changing with temperature so that more Al_2O_3 will be forming as the steel cools until freezing is completed These oxides form as a dispersion throughout the metal phase, and it is inevitable that those oxides formed in the later stages of the process must be trapped in the metal It is thus impossible to make absolutely clean steel by any of the conventional processes which require final deoxidation to prevent the CO evolution. This presentation is highly simplified In any real case, all of the elements present must interact in such a way that the deoxidation products are usually complex associations of the oxides of Fe, Mn, Si, and Al Similarly, these elements all interact with the sulfur in the steel, and this element too is included in the complex The final size, shape, and distribution of these inclusions in the steel is an important factor in its performance in many applications, and a great deal is done in steelmaking to influence these factors

Rimmed steel. This type of steel is selected for good surface characteristics when internal soundness is not critical The ladle treatment for this type is not designed to stop the formation of CO, but rather to control it in order to obtain the following freezing characteristics The molds for this purpose are described as big-end-down and are placed on a heavy bottom section called a stool, so that the mold can be stripped from the solid ingot by lifting the vertical section from the stool No hot top is used The properly deoxidized steel is teemed into the mold, and as freezing begins because of the chilling action of the mold wall, the $C_{(Fe)}$-$O_{(Fe)}$ equilibrium is shifted enough to cause CO evolution at the solid liquid interface

This causes a slight rise of the liquid level in the mold, as well as a stirring of the remaining liquid, which is described as the *rimming* action. This action sweeps any deoxidation products to the top and also prevents the early freezing over of the top surface by maintaining a uniform temperature throughout the liquid. Thus, a *skin of sound*, inclusion-free metal several inches thick is formed. *Finally, the top must freeze over*, and if the deoxidation has been proper, the remaining gas evolution is just right to compensate for the difference in volume between the remaining liquid and solid, and the ingot freezes with a nearly flat top. The *cavity corresponding* to the pipe in killed steel is distributed throughout a large region of the ingot [illegible]

in subsequent rolling operations.

Rimmed steels are cheaper because no hot-topping facilities are required, and yields are greater. They have the best surface for many forming operations because they are inclusion free, and they can be used in many products, such as auto bodies, where an absolutely sound internal structure is not important.

Semikilled steels. As the name implies, these are a compromise between the features of killed and of rimmed products, and the degree of compromise is subject to considerable control and extensive variation. The rimming action can be stopped mechanically by placing a heavy plate over the top of the mold, or chemically by adding Al as additional deoxidizer. Either method is subject to extensive control and variation, which cannot be described in detail here. These types are used for many applications, such as *plate and structural* shapes, where the economies gained are not inconsistent with satisfactory *performance*.

ACID OPEN HEARTH PROCESS

This relatively unimportant process, tonnage-wise, is subject to the restrictions on P and S in the charge which have already been noted for acid processes. It has been used largely for the production of high-quality steels in the higher carbon ranges for steel rolls, castings, and forgings. In these cases, high production rates are unimportant, and the operator pays the necessary price for high-quality scrap and other charge materials. The process is normally operated on an all-solid charge with the necessary mix of ore, pig iron, and scrap to achieve the specified analysis. Although there are many important details related to acid open-hearth steelmaking, the essential features can be deduced from the descriptions already given for acid processes in general, the acid Bessemer process and the principles of deoxidation outlined in connection with the basic open hearth. The normal product is fully deoxidized, can be made with low inclusion content, and is believed to be particularly free from the defects associated with hydrogen.

ELECTRIC FURNACE STEELMAKING

This process began to exert its influence in the industry about 1900 (50 years later than the pneumatic and open-hearth processes, *both of which* depend on an oxidizing atmosphere for their success). The arc furnace develops the necessary temperatures without the requirement of oxygen in the atmosphere, and has, therefore, occupied a unique position as the only process suitable for making many grades of steel with large amounts of such oxidizable alloys as Cr, V, and W. Beginning with the original design of Paul L. T. Héroult, the three-electrode furnace has been developed to the point where it is currently competitive with the open-hearth process for all grades of steel in many situations, and it appears destined for a continued increase in importance. The features of the furnace are shown in Fig. 5, which also indicates the differences in refractories for basic or acid practice. The distinctions between these practices are the same as those described for other processes, and in order to take better advantage of fluctuating sources of raw materials, there is a strong trend in favor of the basic practice. Commercial furnace sizes vary all the way from a few hundred pounds to 200 tons. As a rule, the smaller sizes are used by the castings industry, and the larger sizes are for the production of ingots. *An all-solid charge has been normal*, but experience with the use of hot metal has indicated that this may become *more* prevalent in the near future. The charging of larger furnaces is a critical rate-determining factor, and *it is common to swing the top aside in order to* place the entire charge in the hearth at once from a previously loaded drop bucket.

The making of a hypothetical heat which illustrates the unique *features and versatility* of the *basic* electric-arc process will be described. In most cases, the raw materials for the charge could be selected to eliminate or *minimize* some of the steps included in this example.

Raw materials and reactions. As in the open hearth, the solid charge of ore, limestone, scrap, and pig iron is placed in the hearth after the previous heat has been drained and the bottom repaired. The amounts of these materials are proportioned to achieve the desired specifications when the charge has been melted and brought to temperature, that is, enough oxygen is provided through ore, or mill scale (mainly magnetite, Fe_3O_4), to *refine* the Si, Mn, P, and C present in the charge, and enough lime is included to *give* the slag basicity required to retain the P in the slag. These *items* can be *adjusted*, if necessary, throughout the melting and refining period, and it is now common to supply most of the oxygen requirement directly by lance, rather than by ore or scale. Up to this point, the process and reactions are *quite* comparable to the basic open hearth, with the exception that a relatively small part of the oxygen requirement has come from the furnace *atmosphere*. Air within the furnace is in contact with the graphite elec-

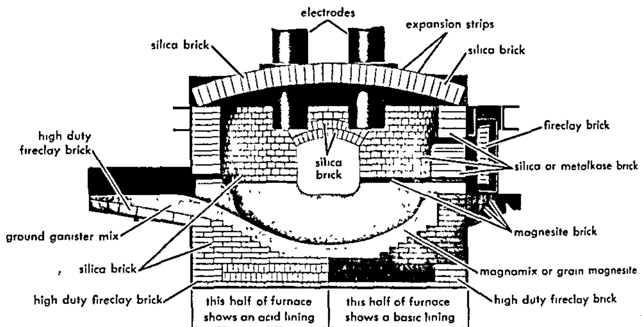


Fig. 5. Section of electric-arc furnace showing refractories for either acid or basic operation. *(Modern Refractory Practice, Harbison-Walker Refractories Company)*

trodes, and all oxygen is rapidly burned to the CO-CO_2 equilibrium for the furnace temperature The basic oxidizing slag will contain the P which has been refined from the charge, so this slag can now be removed from the system by the aid of rabbles, forcing it out the pouring spout or charging door. The furnace can be tilted to assist in this operation. If the charge materials were sufficiently low in P, as is frequently the case, this slag can be left in the furnace to allow the recovery of such valuable alloying metals as Cr which entered the slag during the oxidizing period.

If the oxidizing slag was removed, and the steel specification requires that large amounts of alloy be added, the next step is to make a new, reducing slag by adding burned lime (CaO), C (as crushed electrodes or coke), and flux (CaF_2) as well as small amounts of Al_2O_3 or SiO_2. The exact composition of this slag depends a great deal on the type of alloy being made, but it will be highly basic and often carbidic, that is, contain detectable amounts of CaC_2. In order to be effective, this slag must contain negligible amounts of FeO. In order to achieve this condition rapidly, the steel bath is simultaneously deoxidized in the furnace by addition of ferrosilicon or Al or both, the resulting deoxidation products become a part of the reducing slag. In fact, they are the major source of these oxides in this slag The molten metal bath is thus protected by a nonoxidizing slag, and the atmosphere is kept nonoxidizing by reaction with the electrodes. Power can be supplied to achieve any desired temperature and to melt any amount of alloy which is needed to meet the steel specification. This basic reducing slag performs another important useful function It has a much more favorable chemistry for the removal of sulfur than any other treatment in the entire iron and steelmaking cycle and specifications of the order of 0 002% S can be met

If a P-free oxidizing slag is left on the bath, it is reduced by the same general methods (addition of ferrosilicon, Al, or C), and the corresponding amount of Cr or other alloy reverts to the bath.

In either case, the bath can be sampled and alloys adjusted until they fall within the specified range, and the temperature can be adjusted until the heat is ready to pour Although not as drastic as in the other steelmaking processes, some final ladle deoxidation is still required.

Product steels. The major tonnage steels for which the electric furnace is required are the stainless steels, such as 18–8 (18% Cr, 8% Ni), but all the special high-alloy grades are made by this process and many of the lower-alloy grades with 10% or less of total alloy content, which could be made in the open hearth The latter are made in the electric furnace when they must meet the highest cleanliness for critical service requirements Electric furnace steels are appreciably higher in nitrogen content than the same grades made in the open hearth, and this may be desirable or not according to the application Special procedures are also required to make very low carbon steels (less than about 0 10% C). In general, the electric furnace is the most versatile of all steelmaking processes, since it can be operated as either an acid or basic, oxidizing or reducing process

NEW AND SPECIAL PROCESSES

The production data cited in the table show how the major processes cited dominate the steel in

dustry. The fundamental features of these processes have changed but little since their inception, although there have been continual increases in size of the units and many engineering improvements have led to remarkable modern production rates. It is reasonable to expect no sudden upheaval in the structure of such a basic industry. It is nonetheless important to recognize the current trends in development of these major processes as well as the many special processes which have come into being because of some special requirement which the major processes cannot meet. These developments have occurred at an increasing rate since World War II because a large segment of the European steel industry had a chance to start anew without the restrictions usually imposed by existing equipment, and because commercial oxygen has become available at low cost on a tonnage basis.

Use of oxygen. The use of this oxygen in the conventional steelmaking processes is now well established in a variety of applications. It can be used for combustion in the open hearth to produce higher flame temperatures and develop heat faster or to maintain an existing temperature pattern with lower fuel consumption. When coupled with the higher operating temperatures which are possible with basic roof construction, another well-defined trend, this use of oxygen offers many possible advantages which are yet to be fully exploited.

The use of oxygen for refining by injection into the bath has been important in both the open-hearth and electric furnace. In this application, it may be considered as replacing ore as a source of oxygen with the advantage of more precise control and a more favorable heat balance, resulting in higher production rates. Although a variety of devices can be used for this purpose, the commonest is to insert an iron pipe (about 1 in. in diameter) through the charging door into the bath and to blow oxygen fast enough to cool the pipe and avoid excessive melting of the lance. In the electric furnace, for example, the extra heat obtained with the oxygen lance makes it possible to recover large amounts of Cr from stainless steel scrap.

The most spectacular use of oxygen has been in pneumatic processes where it can enrich the air blown through the tuyeres to refine the bath more rapidly and to develop the heat required to melt more scrap. In the Linnz and Donnewitz (L-D) process the tuyeres have been eliminated, and pure oxygen is blown through a water-cooled nozzle onto the top of the bath in a pear-shaped, basic-lined furnace (Fig. 6). Here the mixing required for rapid refining is obtained from the velocity of the jet, and it appears that this process can realize all of the advantages of a basic slag in refining P and S as well as Si, Mn, C. It can use significantly more scrap than the conventional Thomas process, and it can make a full range of plain C steels with the low nitrogen characteristic of open-hearth steels. Other processes use the oxygen lance with a rotating hearth to obtain mixing of the metal and slag for refining and heat transfer. It seems certain that some of these variations will become increasingly important.

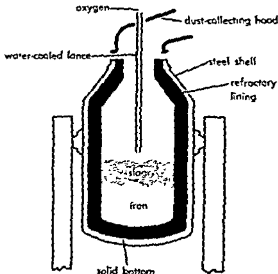


Fig. 6. Schematic section of L-D process vessel. (From *Steel*, 43.27, October, 1958)

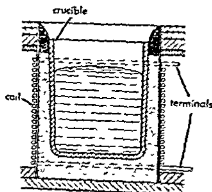


Fig. 7. Section of crucible for high-frequency-induction melting of steel. (From Ajax Electrothermic Corp., Bull. 11-B)

Induction melting. Induction melting has been used to replace the conventional arc furnace in cases where small heats of special alloys are desired, usually for castings. High-frequency power, usually from a motor generator, is supplied to the outer water-cooled coil of the typical furnace shown in Fig. 7. This coil surrounds the crucible containing the charge which acts as the susceptor, and the resulting eddy currents in the charge produce the heat required for melting. This is usually a straightforward melting process in which the proper amounts of high-purity raw materials are placed in the crucible, no slag is formed, and no refining is attempted. Allowance must be made for reactions with air during the meltdown, or a neutral atmosphere provided.

Vacuum processes. All of the processes described thus far require contact of the molten steel with the air or a furnace atmosphere. It has been indicated that these processes necessarily result in a finished steel containing certain amounts of oxy-

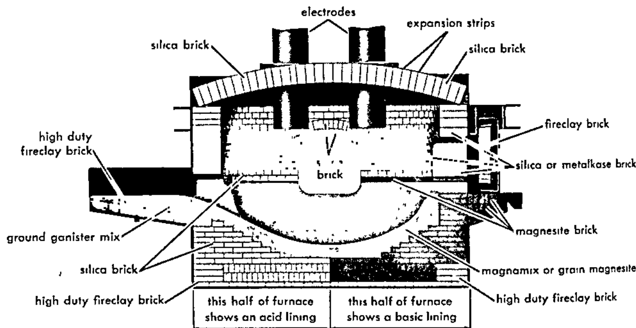


Fig 5. Section of electric-arc furnace showing refractories for either acid or basic operation. (*Modern Refractory Practice, Harbison-Walker Refractories Company*)

trodes, and all oxygen is rapidly burned to the CO-CO_2 equilibrium for the furnace temperature. The basic oxidizing slag will contain the P which has been refined from the charge, so this slag can now be removed from the system by the aid of rabbles, forcing it out the pouring spout or charging door The furnace can be tilted to assist in this operation If the charge materials were sufficiently low in P, as is frequently the case, this slag can be left in the furnace to allow the recovery of such valuable alloying metals as Cr which entered the slag during the oxidizing period

If the oxidizing slag was removed, and the steel specification requires that large amounts of alloy be added, the next step is to make a new, reducing slag by adding burned lime (CaO), C (as crushed electrodes or coke), and flux (CaF_2) as well as small amounts of Al_2O_3 or SiO_2 The exact composition of this slag depends a great deal on the type of alloy being made, but it will be highly basic and often carbidic, that is, contain detectable amounts of CaC_2. In order to be effective, this slag must contain negligible amounts of FeO. In order to achieve this condition rapidly, the steel bath is simultaneously deoxidized in the furnace by addition of ferrosilicon or Al or both; the resulting deoxidation products become a part of the reducing slag. In fact, they are the major source of these oxides in this slag The molten metal bath is thus protected by a nonoxidizing slag, and the atmosphere is kept nonoxidizing by reaction with the electrodes. Power can be supplied to achieve any desired temperature and to melt any amount of alloy which is needed to meet the steel specification. This basic reducing slag performs another important useful function It has a much more favorable chemistry for the removal of sulfur than any other treatment in the entire iron and steelmaking cycle and specifications of the order of 0 002% S can be met.

If a P-free oxidizing slag is left on the bath, it is reduced by the same general methods (addition of ferrosilicon, Al, or C), and the corresponding amount of Cr or other alloy reverts to the bath.

In either case, the bath can be sampled and alloys adjusted until they fall within the specified range, and the temperature can be adjusted until the heat is ready to pour Although not as drastic as in the other steelmaking processes, some final ladle deoxidation is still required.

Product steels. The major tonnage steels for which the electric furnace is required are the stainless steels, such as 18-8 (18% Cr, 8% Ni), but all the special high-alloy grades are made by this process and many of the lower-alloy grades with 10% or less of total alloy content, which could be made in the open hearth The latter are made in the electric furnace when they must meet the highest cleanliness for critical service requirements Electric furnace steels are appreciably higher in nitrogen content than the same grades made in the open hearth, and this may be desirable or not according to the application. Special procedures are also required to make very low carbon steels (less than about 0 10% C). In general, the electric furnace is the most versatile of all steelmaking processes, since it can be operated as either an acid or basic, oxidizing or reducing process

NEW AND SPECIAL PROCESSES

The production data cited in the table show how the major processes cited dominate the steel in-

dustry. The fundamental features of these processes have changed but little since their inception, although there have been continual increases in size of the units, and many engineering improvements have led to remarkable modern production rates. It is reasonable to expect no sudden upheaval in the structure of such a basic industry. It is nonetheless important to recognize the current trends in development of these major processes as well as the many special processes which have come into being because of some special requirement which the major processes cannot meet. These developments have occurred at an increasing rate since World War II because a large segment of the European steel industry had a chance to start anew without the restrictions usually imposed by existing equipment, and because commercial oxygen has become available at low cost on a tonnage basis.

Use of oxygen. The use of this oxygen in the conventional steelmaking processes is now well established in a variety of applications. It can be used for combustion in the open hearth to produce higher flame temperatures and develop heat faster, or to maintain an existing temperature pattern with lower fuel consumption. When coupled with the higher operating temperatures which are possible with basic roof construction, another well-defined trend, this use of oxygen offers many possible advantages which are yet to be fully exploited.

The use of oxygen for refining by injection into the bath has been important in both the open hearth and electric furnace. In this application, it may be considered as replacing ore as a source of oxygen with the advantage of more precise control and a more favorable heat balance resulting in higher production rates. Although a variety of devices can be used for this purpose, the commonest is to insert an iron pipe (about 1 in. in diameter) through the charging door into the bath and to blow oxygen fast enough to cool the pipe and avoid excessive melting of the lance. In the electric furnace for example the extra heat obtained with the oxygen lance makes it possible to recover large amounts of Cr from stainless steel scrap.

The most spectacular use of oxygen has been in pneumatic processes where it can enrich the air blown through the tuyeres to refine the bath more rapidly and to develop the heat required to melt more scrap. In the Linnz and Donnewitz (L-D) process, the tuyeres have been eliminated, and pure oxygen is blown through a water-cooled nozzle onto the top of the bath in a pear-shaped, basic-lined furnace (Fig 6). Here the mixing required for rapid refining is obtained from the velocity of the jet, and it appears that this process can realize all of the advantages of a basic slag in refining P and S as well as Si, Mn, C, it can use significantly more scrap than the conventional Thomas process, and it can make a full range of plain C steels with the low nitrogen characteristic of open-hearth steels. Other processes use the oxygen lance with a rotating hearth to obtain mixing of the metal and slag for refining and heat transfer. It seems certain

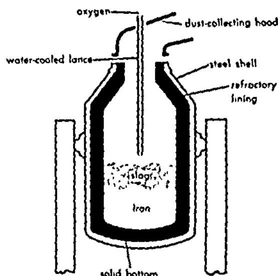


Fig 6 Schematic section of L-D process vessel. *(From Steel, 43 27, October, 1958)*

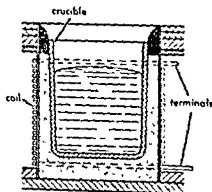


Fig 7 Section of crucible for high frequency induction melting of steel *(From Ajax Electrothermic Corp., Bull. 11-8)*

that some of these variations will become increasingly important.

Induction melting. Induction melting has been used to replace the conventional arc furnace in cases where small heats of special alloys are desired, usually for castings. High-frequency power, usually from a motor generator, is supplied to the outer water-cooled coil of the typical furnace shown in Fig 7. This coil surrounds the crucible containing the charge which acts as the susceptor, and the resulting eddy currents in the charge produce the heat required for melting. This is usually a straightforward melting process in which the proper amounts of high purity raw materials are placed in the crucible, no slag is formed, and no refining is attempted. Allowance must be made for reactions with air during the meltdown, or a neutral atmosphere provided.

Vacuum processes. All of the processes described thus far require contact of the molten steel with the air or a furnace atmosphere. It has been indicated that these processes necessarily result in a finished steel containing certain amounts of oxy-

gen, nitrogen, and hydrogen which in turn must confer special properties, often undesirable, on the product. Special practices such as raw material, slag, and temperature control have been developed to minimize these gas problems, but a more complete solution requires the use of inert atmospheres or vacuum. If the large heat size of an open hearth is required, the most effective current method is to place the mold inside a chamber connected with high-speed pumping equipment and a transfer ladle so the steel is poured through the evacuated space before reaching the mold. This is particularly effective for the removal of hydrogen.

When smaller heats are required, a more effective job can be done by conducting the entire melting and casting procedure within the evacuated space, and this can now be accomplished for heats of 5 tons or less. This can be done by induction melting, in which the evacuated chamber is divided into a series of interlocking sections so that a new solid charge can be introduced without admitting air to the crucible or pouring sections Similarly, ingots or castings can be removed through another vacuum lock. As in normal induction melting, only high-purity raw materials are charged and no ordinary refining is attempted. The vacuum induction-melted product is low in all of the gaseous elements and is the usual starting material for the other type of vacuum melting called the consumable-electrode method. This is also operated in an evacuated chamber which contains a water-cooled copper mold. This holds a small starting block of the alloy to be melted which acts as the cathode for a dc arc The anode is a long section of the same alloy The arc generates enough heat to melt the anode which drips onto the cathode. Proper control of the arc and cooling rate permits a small pool of molten metal at the top of the cathode, and this freezes progressively at the same rate as new melt falls in Thus, a continuous casting is developed which is free of the shrinkage defects described for conventional ingots and is also very low in all gaseous impurities.

These vacuum processes are needed for making many special alloys containing large amounts of easily oxidized metals. They can also be used for the melting of more conventional alloy steel compositions where exceptionally low gas and inclusion

ing, Shaping, and Treating of Steel, 7th ed, 1957.

Steering, power

Manually controlled steering system in which an auxiliary mechanism assists the driver by supplying all or most of the force to steer the road wheels. Principal components of power steering are control valve, power actuator, and source of power. These components operate in conjunction with a hand steering wheel, steering gear, linkages, and steered road wheels (*see* AUTOMOTIVE STEERING). The valve senses any difference between the position of the steering wheel and the corresponding position of the steered wheels and releases power to the actuator until the difference disappears

Application. Self-propelled vehicles have been steered by assistance from servo power for many years. Modern power steering became widespread after 1952 when Chrysler Corporation offered power steering as original equipment on passenger cars. Power steering increases the productive capacity of buses, trucks, and tractors used for transportation, construction, and agriculture by about 20%. On passenger cars, power steering increases safety and convenience by relieving fatigue. Power steering is economical for such diverse applications as heavy duty dump trucks, motor graders, mine cars, and truck cranes. Vehicles that operate over rough terrain, as in timberland towing and earth moving where substantial steering effort is required, benefit especially from powered steering. With power steering, vehicular load can be distributed for compactness and maneuverability.

Servo loop. A servo loop provides the power for steering It is added to a manual steering system together with a control valve The servo loop can be introduced around various parts of the steering

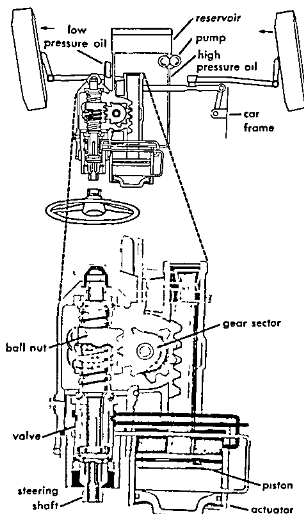


Fig 1. Integral power system operates at steering gear. (*Lincoln*)

system Principal hydraulic power systems, classified by the location of the servo loop, are integral, linkage, in line linkage, and hybrid.

In each system an engine-driven pump with a reservoir delivers hydraulic fluid oil under pressure to a valve. The valve responds to torque from the [illegible] shock from the road

sensitive to very low steering torque so that at all times some manual steering is retained and the driver feels the road

Integral Components of the integral system are combined in a single power steering gear assembly (Fig 1) The steering shaft carries a valve spool and moves slightly along its axis. Springs in the valve assembly float the shaft during normal conditions *Rotation* of the steering wheel causes the

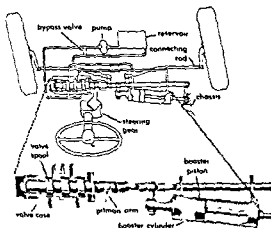


Fig 2 Linkage power steering reacts between chassis and steering linkages. (*Ford*)

shaft to climb up (or down) at the ball nut because the ball nut and the rest of the steering linkage resist any movement The resulting axial motion of the shaft and valve admits high-pressure oil to one side of the actuator piston and allows oil from the other side to return to the reservoir. The piston moves the gear sector and hence the rest of the steering system in a direction to relieve the thrust on the steering shaft, the valve then returns [illegible]

[illegible] actuator or booster cylinder anchors to the chassis Figure 2 shows *one* such arrangement adapted to a cross steering linkage.

The valve consists of a movable spool inside a sealed case The valve is connected through conduits to the high-pressure pump, the low-pressure reservoir and to each side of the booster piston. With the spool in its neutral position, annular grooves on the spool connect oil at low pressure to both sides of the piston The steering system is at rest

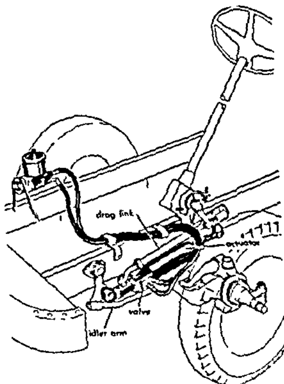


Fig 3 In-line linkages power steering.

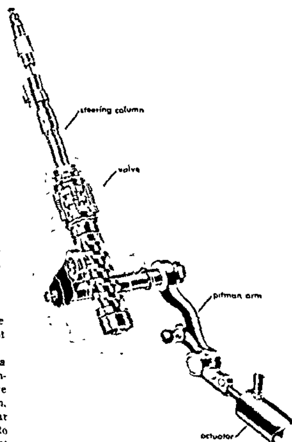


Fig 4 Hybrid power steering.

Steering effort or road reaction deflects the valve spool relative to the valve case. The spool grooves then connect the high-pressure oil line to one or the other side of the booster cylinder so that oil forces the steering linkages in the direction that will return the valve to its neutral position.

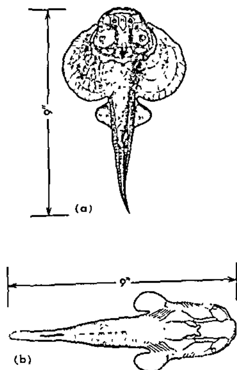
In-line linkage Control valve and actuator are combined in a single assembly for in-line linkage power steering. The system of Fig. 3 connects between drag link and idler arm of a fore-and-aft steering linkage. Motion of the drag link operates a valve similar to that used in other systems. The valve energizes the actuator, which drives the linkage to follow the steering motion or to resist road reaction. When the actuator has reduced the linkage forces, springs in the valve return it to its neutral position until unbalanced forces again set the hydraulic assist into action.

Hybrid. A valve in the manual reserve steering gear and a separate actuator in any other part of the steering system, as in Fig 4, constitute hybrid power steering. Valve and actuator are similar to the corresponding components of other systems and the operation is basically the same.

All four systems are used Each has advantages that suit it to particular vehicular configurations [W.K C.]

Stegoselachii

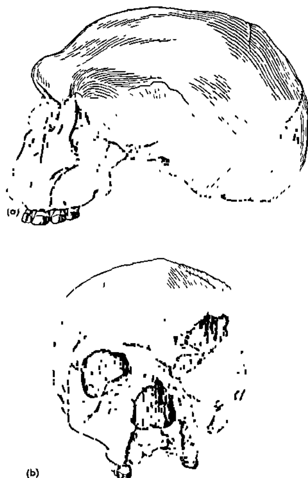
An order of arthrodiran fishes called armored sharks, distinguished by a remarkable reduction of dermal armor including the complete disappearance of pectoral spine. Aside from a very few deeply embedded plates, these raylike fishes display only small, widely spaced skin denticles The Stegoselachii are known principally by fossil specimens from marine and brackish-water sediments of the Lower Devonian of western Europe and North America. The several recognized genera make up two suborders. The first, Stensioellida, is typically represented by *Stensioella*, whose enlarged head and trunk and depressed body habit indicate a bottom-living adaptation. The second suborder, Rhenanida, is represented by *Gemuendina*, a completely benthonic fish, with a flat, broad body, upwardly directed eyes and nostrils, and greatly extended pectoral fins A single incomplete specimen named *Cratoselache*, from the Lower Mississippian of Belgium, possibly represents the last stegoselachian and the youngest known placoderm *See* PLACODERMI. [D H D]



Dorsal views of two stegoselachians from the Lower Devonian. (a) *Gemuendina*; a skatelike form. (b) *Stensioella*, with enlarged head and trunk and depressed body. Originals about 23 cm long (After F. Broili)

Steinheim man

A prehistoric man represented by the damaged skull, without mandible, of a young woman, found in 1933 in a sandpit 12 miles north of Stuttgart, Germany. No artifacts or other human bones were



The Steinheim skull as reconstructed from a cast (a) Right lateral view reversed. (b) Frontal view (From M F. Ashley-Montagu, An Introduction to Physical Anthropology, 2d ed, Charles C Thomas, 1951)

present. The deposit is believed to be late Second Interglacial in date. The skull combines heavy brow ridges and primitive features of the nose region with a relatively small face and a modern-looking braincase. The latter, though low, is vertical-sided, with a moderate forehead and nonprojecting occipital region. The brain volume was approximately 1180 cm³. Relationship has been suggested both with the Neanderthal stock of later date and with the contemporary Swanscombe skull, possibly leading directly to *Homo sapiens* (modern man). The skull is at the Anthropologisches Institut of the University of Tübingen. *See* FOSSIL MAN. [W.W.H.]

Stele

A term including all tissues and regions of plants from the cortex inward: the pericycle, phloem, cambium, xylem and pith (when present). The term was introduced to indicate the unity of structure of the axis (root and stem) of the plant. This axis contains a core (stele) composed of the vascular tissues and associated ground tissues (parenchyma and sclerenchyma) and enclosed by a ground tissue region, the cortex (*see* PARENCHYMA; SCLERENCHYMA). Some anatomists include the endodermis with the stele, others consider it the inner border of the cortex. The steles are classified according to the distribution of the vascular tissues within the ground tissues. This classification refers to the axis in primary state of growth. The three major types of stele are the protostele, siphonostele, and the dictyostele (Fig. 1).

Protostele. This type consists of a solid rod of xylem (no pith) surrounded by phloem (Fig. 2). If the core of xylem has a smooth outline in cross section the protostele is called a haplostele; if the xylem is star-shaped in cross section or if it has ribs radiating from the center, the stele is termed an actinostele, if the xylem is divided into plates the stele is a plectostele. In general, protosteles are found in the stems of the most primitive vascular plants (lycopods, psilophytes, and some ferns) and in the roots of all vascular plants, except those of some dicotyledons and many monocotyledons (*see* FILICALES; LYCOPODIALES; PSILOPHYTALES). From an evolutionary standpoint, the protostele is considered to be the most primitive type of stele. *See* EVOLUTION, ORGANIC.

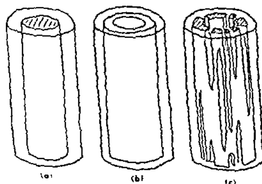


Fig. 1. Diagrams illustrating the types of arrangement of vascular tissues in steles. (a) Protostele. (b) Siphonostele. (c) Dictyostele. (From A. J. Eames and L. H. MacDaniels, Introduction to Plant Anatomy, 2d ed., McGraw-Hill, 1947)

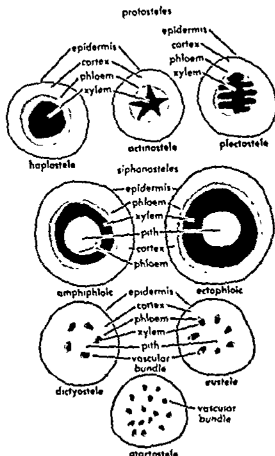


Fig. 2. Types of stele.

Siphonostele. There are two types of stele with pith: the amphiphloic siphonostele, or solenostele, with phloem on both the inside and outside of the xylem; and the ectophloic siphonostele, with phloem only on the outside of the xylem. The former is found in the stems of some ferns, and the latter, in the stems of some ferns, gymnosperms, and angiosperms (*see* ANGIOSPERMAE; GYMNOSPERMAE). Siphonosteles are considered to represent a more advanced evolutionary state than protosteles.

Dictyostele. This is a type of siphonostele found in fern stems in which the vascular tissue is divided by many gaps, or parenchymal regions, to form a network, and hence the xylem and phloem are arranged in separate concentric vascular bundles, as seen in cross section (*see* VASCULAR BUNDLES). A similar type of siphonostele, found in most ...

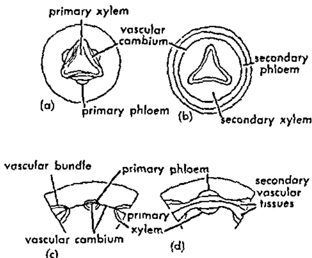


Fig. 3. (a,b) Diagrams of cross sections of a root in two stages of secondary growth. The root has a protostele. (c,d) Diagrams of cross sections of parts of stem in secondary state of growth. The stem has a eustele in primary state of growth, a continuous cylinder of vascular tissues in secondary state of growth. (From A. J. Eames and L. H. MacDaniels, Introduction to Plant Anatomy, 2d ed., McGraw-Hill, 1947)

tionary derivative of the dissected siphonostele, the atactostele, is the type found in the monocotyledons in which the vascular bundles do not occur in a ring but are dispersed throughout the center of the stem.

Secondary growth in steles. In gymnosperms and dicotyledons the vascular cambium commonly produces secondary phloem and secondary xylem. In the root, the secondary vascular tissues completely imbed the core of primary xylem, whereas the primary phloem is pushed outward and more or less completely crushed. In the stems, secondary vascular tissues are often formed between the xylem and phloem and also in the interfascicular regions, that is, between the bundles. Thus, in contrast to the eustelic condition in the primary state (network of bundles), a continuous cylinder of vascular tissues is formed in the secondary state (Fig. 3). *See* ENDODERMIS; MERISTEM, LATERAL; PERICYCLE; PHLOEM; PITH; XYLEM; *see also* ROOT (BOTANY); STEM (BOTANY). [O.T.]

Bibliography: *See* PLANT ANATOMY.

Stellar evolution

The changes that take place in a star. Even though man's knowledge of stars is largely indirect and theoretical, it rests on well-tested physical laws, which link the unobservable properties of stellar interiors to quantities that are susceptible to more or less direct measurements. The most important of these quantities are (1) stellar luminosity (or absolute magnitude), which can be inferred from the apparent magnitude of a star if its distance and the effect of interstellar absorption are known; (2) stellar color (or effective temperature), which is determined as a correction for interstellar reddening; (3) stellar mass, which can be determined for some, but not all, stars that belong to binary systems; and (4) chemical composition in the outer layers, which can be inferred partially and with considerable uncertainty from an analysis of the stellar spectrum. This article summarizes present knowledge of stellar structure and evolution and presents the principal physical arguments on which that knowledge rests.

Stellar structure. A nonrotating, chemically homogeneous mass of gas held together by its own gravitation provides the simplest imaginable model of a star, but one that represents fairly well the observable properties of a large and well-defined class of stars, specifically those belonging to the main sequence of the Hertzsprung-Russell (H-R) diagram (*see* STAR). Theory and observation agree that main-sequence stars consist largely of hydrogen (70–90% by weight) and helium (10–30%). The percentage of heavier elements varies from about 3% in Population I stars to an undetermined lower limit which may be considerably less than 1% in Population II stars. The helium-to-hydrogen ratio does not strongly influence the color or luminosity of a star, as long as it is not too large, and it is therefore difficult to determine theoretically. Unfortunately it is also difficult to determine observationally, because helium lines occur chiefly in the spectra of very hot stars whose outer layers, where the lines are formed, are not yet well understood. The abundance of heavy elements, on the other hand, has an important effect on the color and luminosity of a star, and it is easier to determine spectroscopically than the helium abundance.

Stars that do not belong to the main sequence fall into two main groups: red giants, whose chief peculiarities result from the fact that they possess much more extended atmospheres than main-sequence stars; and white dwarfs, which are exceedingly faint for their color (as compared with main-sequence stars) and have astonishingly high mean densities, of the order of 10^6 g/ml. Modern theoretical work indicates that red giants and white dwarfs began their stellar careers as normal main-sequence stars and owe their distinctive properties to evolutionary changes incident on the exhaustion of nuclear energy sources. Red giants are now believed to be stars not much more massive than the Sun but markedly inhomogeneous in chemical composition. White dwarfs are chemically homogeneous, or nearly so, but differ radically from main-sequence stars in several fundamental ways; hence they will be discussed separately.

The discussion turns now to specific motions and problems in the theory of stellar structure and evolution.

Temperature, density, pressure. Typical values for temperature T, density ρ, and pressure P in main-sequence stars are $T = 10^7$ °K, $\rho = 1$ g/ml, $P = 2 \times 10^{16}$ dyne/cm²; at the center of the Sun $T = 1.5 \times 10^7$ °K, $\rho = 100$ g/ml, $P = 3 \times 10^{17}$ dyne/cm². Even at this high density (the normal density of water is 1 g/ml) stellar material remains a perfect gas, because the enhanced thermal motions of the ions and electrons more

Great Nebula in Orion (above), visible as the middle "star" in the sword of Orion, is a gas cloud excited to incandescence by hot stars in its center. Stars may form in such a region. Crab Nebula in Taurus (below) is debris from a stellar explosion seen as a nova by Oriental astronomers in 1054. High-energy electrons cause the gas to glow. These colors are not visible because of low intensity; even sensitive color film requires exposures of several hours through 200-inch telescope. *(Photographs by William Miller, research photographer, California Institute of Technology)*

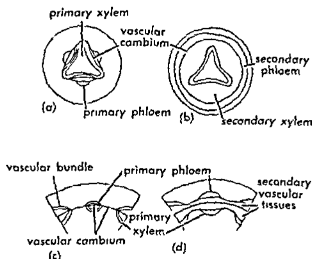


Fig. 3 (a,b) Diagrams of cross sections of a root in two stages of secondary growth. The root has a protostele. (c,d) Diagrams of cross sections of parts of stem in secondary state of growth. The stem has a eustele in primary state of growth, a continuous cylinder of vascular tissues in secondary state of growth. (From A. J. Eames and L. H. MacDaniels, *Introduction to Plant Anatomy*, 2d ed., McGraw-Hill, 1947)

tionary derivative of the dissected siphonostele, the atactostele, is the type found in the monocotyledons in which the vascular bundles do not occur in a ring but are dispersed throughout the center of the stem

Secondary growth in steles. In gymnosperms and dicotyledons the vascular cambium commonly produces secondary phloem and secondary xylem. In the root, the secondary vascular tissues completely imbed the core of primary xylem, whereas the primary phloem is pushed outward and more or less completely crushed. In the stems, secondary vascular tissues are often formed between the xylem and phloem and also in the interfascicular regions, that is, between the bundles. Thus, in contrast to the eustelic condition in the primary state (network of bundles), a continuous cylinder of vascular tissues is formed in the secondary state (Fig. 3). *See* ENDODERMIS; MERISTEM, LATERAL; PERICYCLE; PHLOEM, PITH, XYLEM; *see also* ROOT (BOTANY), STEM (BOTANY). [O.T.]

Bibliography: *See* PLANT ANATOMY.

Stellar evolution

The changes that take place in a star. Even though man's knowledge of stars is largely indirect and theoretical, it rests on well-tested physical laws, which link the unobservable properties of stellar interiors to quantities that are susceptible to more or less direct measurements. The most important of these quantities are (1) stellar luminosity (or absolute magnitude), which can be inferred from the apparent magnitude of a star if its distance and the effect of interstellar absorption are known, (2) stellar color (or effective temperature), which is determined as a correction for interstellar reddening; (3) stellar mass, which can be determined for some, but not all, stars that belong to binary systems; and (4) chemical composition in the outer layers, which can be inferred partially and with considerable uncertainty from an analysis of the stellar spectrum. This article summarizes present knowledge of stellar structure and evolution and presents the principal physical arguments on which that knowledge rests.

Stellar structure. A nonrotating, chemically homogeneous mass of gas held together by its own gravitation provides the simplest imaginable model of a star, but one that represents fairly well the observable properties of a large and well-defined class of stars, specifically those belonging to the main sequence of the Hertzsprung-Russell (H-R) diagram (*see* STAR). Theory and observation agree *that main-sequence stars consist largely of hydrogen* (70–90% by weight) and helium (10–30%). The percentage of heavier elements varies from about 3% in Population I stars to an undetermined lower limit which may be considerably less than 1% in Population II stars. The helium-to-hydrogen ratio does not strongly influence the color or luminosity of a star, as long as it is not too large, and it is therefore difficult to determine theoretically. Unfortunately, it is also difficult to determine observationally, because helium lines occur chiefly in the spectra of very hot stars whose outer layers, where the lines are formed, are not yet well understood. The abundance of heavy elements, on the other hand, has an important effect on the color and luminosity of a star, and it is easier to determine spectroscopically than the helium abundance.

Stars that do not belong to the main sequence fall into two main groups: red giants, whose chief peculiarities result from the fact that they possess much more extended atmospheres than main-sequence stars, and white dwarfs, which are exceedingly faint for their color (as compared with main-sequence stars) and have astonishingly high mean densities, of the order of 10^6 g/ml. Modern theoretical work indicates that red giants and white dwarfs began their stellar careers as normal main-sequence stars and owe their distinctive properties to evolutionary changes incident on the exhaustion of nuclear energy sources. Red giants are now believed to be stars not much more massive than the Sun but markedly inhomogeneous in chemical composition. White dwarfs are chemically homogeneous, or nearly so, but differ radically from main-sequence stars in several fundamental ways; hence they will be discussed separately.

The discussion turns now to specific motions and problems in the theory of stellar structure and evolution

Temperature, density, pressure. Typical values for temperature T, density ρ, and pressure P in main-sequence stars are $T = 10^6$ °K, $\rho = 1$ g/ml, $P = 2 \times 10^{14}$ dyne/cm², at the center of the Sun $T = 1.5 \times 10^7$ °K, $\rho = 100$ g/ml, $P = 3 \times 10^{17}$ dyne/cm². Even at this high density (the normal density of water is 1 g/ml), stellar material remains a perfect gas, because the enhanced thermal motions of the ions and electrons more

Great Nebula in Orion (above), visible as the middle "star" in the sword of Orion, is a gas cloud excited to incandescence by hot stars in its center. Stars may form in such a region. Crab Nebula in Taurus (below) is debris from a stellar explosion seen as a nova by Oriental astronomers in 1054. High-energy electrons cause the gas to glow. These colors are not visible because of low intensity; even sensitive color film requires exposures of several hours through 200-inch telescope. *(Photographs by William Miller, research photographer, California Institute of Technology)*

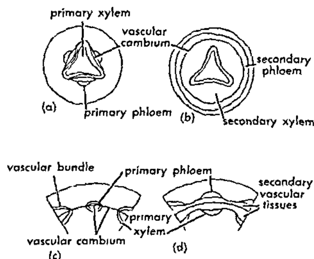


Fig. 3. (a,b) Diagrams of cross sections of a root in two stages of secondary growth. The root has a protostele. (c,d) Diagrams of cross sections of parts of stem in secondary state of growth. The stem has a eustele in primary state of growth, a continuous cylinder of vascular tissues in secondary state of growth. (From A. J. Eames and L. H. MacDaniels, Introduction to Plant Anatomy, 2d ed., McGraw-Hill, 1947)

tionary derivative of the dissected siphonostele, the atactostele, is the type found in the monocotyledons in which the vascular bundles do not occur in a ring but are dispersed throughout the center of the stem.

Secondary growth in steles. In gymnosperms and dicotyledons the vascular cambium commonly produces secondary phloem and secondary xylem. In the root, the secondary vascular tissues completely imbed the core of primary xylem, whereas the primary phloem is pushed outward and more or less completely crushed. In the stems, secondary vascular tissues are often formed between the xylem and phloem and also in the interfascicular regions, that is, between the bundles. Thus, in contrast to the eustelic condition in the primary state (network of bundles), a continuous cylinder of vascular tissues is formed in the secondary state (Fig. 3). *See* ENDODERMIS; MERISTEM, LATERAL; PERICYCLE; PHLOEM; PITH; XYLEM; *see also* ROOT (BOTANY); STEM (BOTANY). [O.T.]

Bibliography. *See* PLANT ANATOMY.

Stellar evolution

The changes that take place in a star. Even though man's knowledge of stars is largely indirect and theoretical, it rests on well-tested physical laws, which link the unobservable properties of stellar interiors to quantities that are susceptible to more or less direct measurements. The most important of these quantities are (1) stellar luminosity (or absolute magnitude), which can be inferred from the apparent magnitude of a star if its distance and the effect of interstellar absorption are known; (2) stellar color (or effective temperature), which is determined as a correction for interstellar reddening; (3) stellar mass, which can be determined for some, but not all, stars that belong to binary systems; and (4) chemical composition in the outer layers, which can be inferred partially and with considerable uncertainty from an analysis of the stellar spectrum. This article summarizes present knowledge of stellar structure and evolution and presents the principal physical arguments on which that knowledge rests.

Stellar structure. A nonrotating, chemically homogeneous mass of gas held together by its own gravitation provides the simplest imaginable model of a star, but one that represents fairly well the observable properties of a large and well defined class of stars, specifically those belonging to the main sequence of the Hertzsprung-Russell (H-R) diagram (*see* STAR). Theory and observation agree that main-sequence stars consist largely of hydrogen (70–90% by weight) and helium (10–30%). The percentage of heavier elements varies from about 3% in Population I stars to an undetermined lower limit which may be considerably less than 1% in Population II stars. The helium-to-hydrogen ratio does not strongly influence the color or luminosity of a star, as long as it is not too large, and it is therefore difficult to determine theoretically. Unfortunately, it is also difficult to determine observationally, because helium lines occur chiefly in the spectra of very hot stars whose outer layers, where the lines are formed, are not yet well understood. The abundance of heavy elements, on the other hand, has an important effect on the color and luminosity of a star, and it is easier to determine spectroscopically than the helium abundance.

Stars that do not belong to the main sequence fall into two main groups: red giants, whose chief peculiarities result from the fact that they possess much more extended atmospheres than main-sequence stars; and white dwarfs, which are exceedingly faint for their color (as compared with main-sequence stars) and have astonishingly high mean densities, of the order of 10^6 g/ml. Modern theoretical work indicates that red giants and white dwarfs began their stellar careers as normal main-sequence stars and owe their distinctive properties to evolutionary changes incident on the exhaustion of nuclear energy sources. Red giants are now believed to be stars not much more massive than the Sun but markedly inhomogeneous in chemical composition. White dwarfs are chemically homogeneous, or nearly so, but differ radically from main-sequence stars in several fundamental ways; hence they will be discussed separately.

The discussion turns now to specific motions and problems in the theory of stellar structure and evolution.

Temperature, density, pressure. Typical values for temperature T, density ρ, and pressure P in main-sequence stars are $T = 10^6$ °K, $\rho = 1$ g/ml, $P = 2 \times 10^{14}$ dyne/cm²; at the center of the Sun $T = 1.5 \times 10^7$ °K, $\rho = 100$ g/ml, $P = 3 \times 10^{17}$ dyne/cm². Even at this high density (the normal density of water is 1 g/ml) stellar material remains a perfect gas, because the enhanced thermal motions of the ions and electrons more

than compensate for the increased interaction that results from their crowding. Significant departures from the perfect-gas law are first encountered, in stellar interiors, only at the much higher densities that prevail in white dwarfs and the cores of red giants.

For a star to be in mechanical equilibrium, the pressure force acting on any spherical surface concentric with the stellar surface must just balance the weight of all the overlying material. In particular, the pressure at the center of a star must balance its entire weight. This is proportional to the square of the star's mass, which accounts for the dominant role played by gravitation in masses of stellar size (the mass of the Sun is 2×10^{33} g) and for the pressures, enormous by terrestrial standards, that prevail in stellar interiors. The central pressure is related to mass M and radius R by the approximate equation $P = K\, GM^2/R^4$, where G is Newton's gravitational constant 6.67×10^{-8} cm^3/(g)(sec^2) and constant K is of the order of unity.

The pressure inside a star comes partly from the thermal motions of the electrons and ions (the ordinary gas pressure) and partly from the pressure exerted by the radiation field. A minute effect in ordinary terrestrial experience, radiation pressure becomes important in stellar interiors because it increases as the fourth power of temperature; it is 10^{16} times as great at 10^{7}°K as at 10^{3}°K. The ratio between radiation pressure and gas pressure increases with increasing luminosity along the main sequence, but attains appreciable values only in the luminous blue supergiants. Theoretically, a star in which radiation pressure predominated would be unstable.

Heat flow. The condition of mechanical equilibrium and the perfect-gas law supply two relations among pressure, density, and temperature. A third relation is supplied by the condition of thermal equilibrium: every element of stellar material must emit and absorb heat at equal rates, otherwise some regions would gradually become hotter or cooler.

The implications of this requirement depend on the type of heat flow. There are three possibilities: conduction, radiative transfer, and convection. Except in white dwarfs conduction is much less efficient than the other two mechanisms and may be ignored. In all stars except white dwarfs radiative transfer accounts for most of the heat flow, but in some stars convection also plays a role.

Radiative transfer. Radiative transfer results from the [illegible] material re[illegible] from relat[illegible] ones but [illegible] cenes. Th[illegible] in a net flow of radiant energy from hot regions to cool ones, which is radially outward. As the radiation makes its way from the center to the surface it gradually changes in quality, becoming redder and redder as it is absorbed and reemitted by stellar material at lower and lower temperatures. The radiation that finally emerges at the surface is characterized by a much lower temperature, called the effective temperature of the star, than that of its source, the deep interior. The effective temperature of the Sun is about 5800°K.

Convection. Unlike radiative heat transfer and conduction, convection does not automatically result from a nonuniformity of temperature, because it entails fluid motions (convection currents), which can occur in a closed system only if the motions tend to reduce the potential energy of the system. This will be the case in a star if any fluid element that is initially less dense than the surrounding material remains so as it falls and contracts. Because a rising or falling fluid element is always at the same pressure as the surrounding material, it will be less dense than its surroundings if its temperature is higher, more dense if its temperature is lower. Thus the important criterion for convection in stellar interiors: Convection prevails if, and only if, it tends to establish a temperature gradient that is less steep than the one which would be established in its absence.

Because the convective gradient is normally steeper than the radiative one in stellar interiors, radiative transfer is the normal mode of heat flow. But there are two situations in which convection predominates: in cores of upper main-sequence stars, and in hydrogen convection zones. In the central cores of stars that derive their energy from the carbon proton capture cycle, such as in upper-main-sequence stars, the temperature gradient that would be set up in the absence of convection is exceptionally steep as a result of the extreme temperature sensitivity of the nuclear reaction rate (see THERMONUCLEAR REACTION). These stars have convective cores, which increase in relative size with increasing luminosity, attaining a radius one-fifth the stellar radius in a main-sequence star of 10 solar masses. Lower main-sequence stars, which derive their energy from the proton-proton chain have radiative cores.

The convective gradient is exceptionally flat in a partially ionized gas, because the addition or withdrawal of heat tends to alter the degree of ionization rather than the temperature. In a star, the region where hydrogen is partially ionized, if such a region exists, ought to be convective. This region is known as the hydrogen convection zone. In the Sun it is thought to begin nine tenths of the way out from the center and to extend nearly, but not quite, to the surface, being surmounted by a thin layer in radiative equilibrium. Upper main-sequence stars do not possess a hydrogen convection zone because all their hydrogen is fully ionized. Lower-main-sequence stars, on the other hand, possess deep convective envelopes. In a star six tenths as massive as the Sun, convection sets in two-thirds of the way out and persists to the surface. Thus, the most extensive convection regions occur in stars at the extremes of the main sequence; the most luminous stars have convective cores surmounted by a radiative envelope, and the least luminous have

that lower-main-sequence stars may contract in much shorter intervals.

When the *central temperature* and density of a contracting star attain sufficiently high values, thermonuclear reactions begin to supply some of the energy, and the contraction slows down. Finally the reaction rate just balances the radiation rate, and contraction ceases. A slight over-contraction would cause the reaction rates to exceed the radiation rate, the star would expand, and the balance would be restored. Thus the equilibrium is stable.

Main-sequence stars. As a star contracts, both its effective temperature and its luminosity increase, temperature more rapidly than luminosity. In the subsequent phase of stellar evolution observable properties of a star change at a much slower rate, and its representative point in the H-R diagram stays *on or close* to the main sequence. The duration of this phase ranges from a few million years for the most luminous upper-main-sequence stars to upwards of 10^{11} years for lower-main-sequence stars. During this period, hydrogen in the core is being *converted to helium*. The main-sequence phase ends when this process has been completed and the core consists wholly of helium and heavy elements.

The evolution of a hydrogen-burning star depends critically on what happens to the helium that is produced. No appreciable mixing occurs between the hydrogen-burning core and the rest of the star, because this *core* is surrounded by a quiescent zone in radiative equilibrium which effectively prevents intermixing of the material on either *side of it*. Stellar rotation is unable to *promote* internal convection and hence mixing. Consequently the helium *produced* in the core remains there. Because *no* helium is produced in the outer layers, *the star gradually develops* a chemically inhomogeneous structure. This inhomogeneity is of *crucial importance* for the subsequent evolution. As the *difference in* molecular weight between *core* and envelope becomes more marked the *envelope must distend* to preserve *equilibrium*. This tends to lower the star's effective temperature. Also, the *development* of a helium core enhances the luminosity, which tends to raise the eff-

th

lu ... the total change in *color* and luminosity is, however, small during this phase.

In stars with *quiescent* (radiative) cores, the region where hydrogen is exhausted gradually increases in size, meanwhile the *zone* where nuclear reactions are continuing steadily grows smaller. In stars with convective cores, the core remains chemically homogeneous, but its hydrogen content gradually diminishes as the helium *that is produced* at the center is transported by convection to the outer part of the core and the hydrogen in the *outer* part is carried to the center. In either case an inert isothermal core ultimately results, made up largely of helium and containing about 10% of the star's mass. In a thin shell surrounding the *core* hydrogen is still being consumed. The extended envelope is not sufficiently hot and dense to support thermonuclear reactions.

Red giants. The development of a helium core results in an unstable configuration. The core contracts, growing hotter and denser, while the envelope expands to perhaps a hundred times its initial size: the star is now a red giant. The theory of red giants is still in a formative stage, but it seems likely that there are two distinct evolutionary paths that a star can follow and that the choice depends on the stellar mass. If this is below a certain value, not much greater than the mass of the Sun, the contraction of the core soon results in a stable configuration. The stability is due to the onset of degeneracy in the core, *which increases the pressure* to such an extent that the contraction virtually ceases. The hydrogen burning shell gradually works its way outwards, the star *meanwhile becoming progressively brighter and redder.*

Because *red giants are so distended*, their outermost layers are weakly bound by the gravitational attraction of the underlying matter. A considerable portion of this weakly held envelope may *subsequently* be ejected into space. Observational evidence for mass ejection by luminous red giants has been found in recent years, but *the quantity of matter* ejected and the times of ejection *are not* yet known.

In stars considerably more massive than the Sun, the *onset of degeneracy* in the contracting helium core is *not sufficient* to halt the contraction, which continues until the temperature in the core has *risen to the point* where helium-burning reactions begin to occur at a *significant* rate. These reactions require exceedingly high densities and *temperatures in the region of* 10^{8}°K (see FUSION, NUCLEAR). The react ...

oxygen, neon, ...

following the ...

higher temperatures and densities and to a *series* of intricate nuclear reactions in which the production and capture of neutrons is thought to play an *important part*. These reactions release energy so *rapidly* that the star may *explode*, thus becoming a supernova (*see* NOVA).

White dwarfs. Supernovae *are* thought to be the end *product* of one of two possible lines of evolution, the line *taken by stars* considerably more massive than the Sun. The other *line* of evolution leads to white dwarfs. Evolution of a star from a *red* giant to a white *dwarf* is not understood in detail. The extended hydrogen envelope of the red giant is partly converted to helium and added to the core and partly, perhaps, ejected. Ultimately there results a star that is *all core* and that is neither burning nuclear fuel nor contracting but simply growing progressively cooler.

Except in their outermost layers white dwarfs are in a state of al*most* complete degeneracy. This means that the pressure forces which k... from a ...

a

radiative cores surmounted by a convective envelope.

Convection does not interfere with the radiative transport of energy, except insofar as it reduces the temperature gradient. Thus, even in a convection zone, radiative transport may account for a substantial proportion of the heat flow.

Deep interior. Even in relatively cool stars the region where hydrogen and helium are completely ionized contains virtually all the mass. In the deep interior, the molecular weight lies between ½, the value appropriate for hydrogen alone, and ⅘, the value appropriate for helium alone. As hydrogen is converted to helium by thermonuclear reactions, the molecular weight gradually increases and tends to enhance the luminosity.

At the pressures and temperatures that prevail in stellar interiors the heavy elements are highly, but not completely, ionized. Despite their comparative paucity, partially stripped heavy atoms contribute strongly to the opacity of stellar material, and for this reason the heavy-element abundance is an important parameter in the theory of stellar evolution.

Opacity plays a role in radiative transfer which is analogous to that of resistance in molecular conduction. If the opacity of a star could be everywhere doubled, its luminosity would be halved. The *reciprocal of the opacity is the distance that a* quantum of radiation, or photon, can be expected to travel before being absorbed or scattered. A typical value for this distance in stellar interiors is 10^{-3} cm. Because the free path of a photon is so small compared with stellar radii (the radius of the Sun is 7×10^{10} cm) and because radiative transfer is a diffusion process, it takes a long time, of the order of 10^{6} years, for any change that occurs in the deep interior to affect the luminosity. Moreover, sharp changes in the deep interior undergo a *considerable smoothing before they show up as* changes in luminosity. Therefore, variations in luminosity with periods of the order of days or fractions of a day, such as occur among variable stars, cannot be attributed to variations in the rate of energy production.

Stellar opacity has three main sources, all of which are photon-electron interactions: (1) photoionization, (2) free-free transitions, and (3) electron (Thompson) scattering. All contribute to the specific opacity (opacity per gram of stellar material). Contributions of bound-free and free-free transitions increase with density and decrease with increasing temperature. Electron scattering makes the dominant contribution in the very hot interiors of upper-main-sequence stars and red giants; bound-free and free-free transitions, in middle- and lower-main-sequence stars.

The contribution of bound-free transitions to the opacity is proportional to the abundance of heavy elements, because all the hydrogen and helium is fully ionized in the deep interior, and is approximately equal to the free-free contribution when the abundance of heavy elements is about 1%. Hence the bound-free contribution is dominant in Population I (disk-type) stars, whose heavy element abundances may reach 3%, and insignificant in Population II (halo-type) stars, which are believed to have heavy-element abundances considerably less than 1%. Halo-type stars should therefore be brighter and bluer than disk-type stars of the same mass if the radii are nearly the same.

Sources of stellar energy. There are strong reasons for believing that the Sun has been shining at essentially its present rate (4×10^{33} ergs/sec) for at least 5×10^{9} years, so that a total of about 6×10^{50} ergs has already been radiated. This implies an average energy yield of 3×10^{17} ergs/g, which is several orders of magnitude higher than could be provided by chemical reactions, which have yields of the order of 10^{12} ergs/g.

A more promising source of stellar energy is gravitational contraction. About half the energy released when a star contracts under its own gravitation is used to increase the internal temperature; the rest is radiated away. Had the Sun initially been much more distended than it is now, about 10^{14} ergs/g could have been released by gravitational contraction, but this is only enough to keep the Sun shining at its present rate for a few million years.

Thermonuclear reactions are the main source of stellar energy (*see* CARBON-NITROGEN CYCLE; PROTON-PROTON CHAIN). In upper-main-sequence stars, the carbon cycle is the main reaction; in the Sun and less massive main-sequence stars, the proton-proton chain provides the energy. Both these reactions convert hydrogen to helium and yield about 6×10^{18} ergs/g. This is enough to keep the Sun and less massive main-sequence stars shining for 5×10^{9} years and more, but upper-main-sequence stars would exhaust their supplies of hydrogen in much shorter periods of time because luminosity *increases much more rapidly than mass along the* main sequence. A star of 10 solar masses, for instance, is perhaps 10^{4} times as luminous as the Sun and would therefore exhaust its hydrogen a thousand times as fast as the Sun. The upper-main-sequence stars that we observe today must, therefore, have been formed at a comparatively recent epoch in the history of the Galaxy.

Contraction of protostars. Little is known about the properties of newly formed stars, but it seems likely that the majority of stars begin as cool, distended, irregular blobs of gas (*see* COSMOGONY). Being too cool and too tenuous to support thermonuclear reactions, a protostar must contract to supply the energy that it radiates, and the rate of radiation determines the rate of contraction. A star that is in this stage of its evolution will appear considerably redder and somewhat less luminous than a main-sequence star of the same mass. Clusters of such stars have been found, and most of these contain upper-main-sequence stars as well, a sign of extreme youth.

According to current theoretical views, the contraction of upper-main-sequence stars continues for about 10^{5} years. There is some evidence, however,

and still others are so short and inconspicuous that the plants are said to be stemless, or acaulescent (Fig 1). When stems lie flattened immediately above but not on the ground, with tips curved upward, they are said to be decumbent, as in juniper. If stems lie flat on the ground but do not root at the nodes (joints), the stem is called procumbent or prostrate, as in purslane. If a stem creeps along the ground, rooting at the nodes, it is said to be repent or creeping, as in ground ivy.

Shape and texture. Most stems are cylindrical and tapering (terete), appearing circular in cross section; others may be quadrangular, as in mints; or triangular, as in some sedges.

Herbaceous stems (annuals and herbaceous perennials) die to the ground after blooming or at the end of the growing season (*see* ANNUAL PLANTS). They usually contain little woody tissue Woody stems (perennials) have considerable woody supporting tissue and live from year to year (*see* PERENNIAL PLANTS) A woody plant with no main stem or trunk, but usually with several stems developed from a common base at or near the ground, is known as a shrub Suffrutescent stems are intermediate between herbaceous and shrubby and become partly

may be thickened, as in Solomon's seal, slender, as in *Bermuda* grass; or contracted at regular intervals as in pepperroot, in which case it is a *moniliform* rhizome. A caudex is an upright *perennial* underground stem, it is much like a rhizome but

as in white potato A corm is a short, erect, fleshy subterranean stem usually broader than high and often covered with dry membranous coats, as in crocus A bulb may be regarded as a short subterranean stem with many overlapping fleshy leaf bases or scales

a support to which the plant is then attached as in grape A cladophyll, or phylloclade is a usually flattened stem resembling a leaf that arises in the axil of a minute leaf (scale), as in asparagus In some instances the resemblance of a cladophyll to a leaf is noted only or mainly in its green color, as in certain species of cacti *See* PLANT ORGANS

[N.A.]

EXTERNAL FEATURES OF STEMS

Morphologically, a branch or shoot consists of a stem, or axis, and leafy appendages The stem may be distinguished from other plant parts by certain external features True stems arise from the

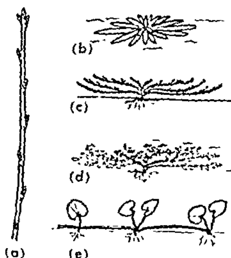


Fig. 1. Kinds of stems. (a) Woody (plum). (b) Acaulescent (evening primrose) (c) Decumbent (juniper). (d) Procumbent (purslane). (e) Repent (ground ivy).

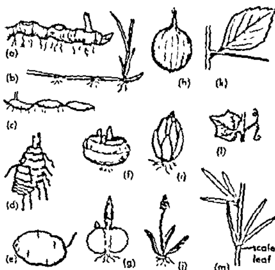


Fig 2 Specialized stems. (a) Rhizome (Solomon's seal). (b) Rhizome (grass) (c) Moniliform rhizome (pepperroot) (d) Caudex (trillium) (e) Tuber (potato) (f) Corm (crocus) (g) Section of corm (h) Tunicated bulb (onion). (i) Scaly bulb (lily) (j) Scape (English plantain) (k) Thorn (hawthorn) (l) Tendril (grape). (m) Cladophyll (asparagus).

epicotyl of the seed or from buds, have nodes and internodes, bear leaves and buds (and some-

the division of the stem into nodes and internodes and the structural relations between stems and leaves. *See* LEAF (BOTANY), SEED (BOTANY).

Nodes. The nodes are the regions of the primary stem or axis where leaves and axillary or accessory buds arise. The number of leaves at a node is usually specific for each kind of plant In deciduous

agitation of the electrons and ions, but rather from motions of the electrons only (the ionic motions contribute virtually nothing to the pressure) that are nonthermal in origin and persist even at the absolute zero of temperature. These nonthermal motions are a consequence of the celebrated uncertainty principle, one of the basic laws of quantum mechanics, which specifies, among other things, that the momentum of a particle known to be within a certain volume cannot be reduced below a value that depends on this volume, and it increases as the volume decreases (*see* UNCERTAINTY PRINCIPLE). An increase in the density of a gas diminishes the volume per particle and thus increases the zero-point pressure, as the nonthermal part of the pressure is called

As a star contracts and the material of which it is composed approaches degeneracy, gravitational forces and zero-point pressure forces increase. At first the latter increase faster than the former, but at high densities the two rates become equal. If the mass of the star is less than a critical value, which depends on the molecular weight but is in any case not much greater than the mass of the Sun, there is a unique value of density (and hence of stellar radius) at which the pressure forces just balance the gravitational ones; this is the equilibrium configuration of a white dwarf. The greater the mass of the star (provided that it is less than the critical value) the smaller is its radius If the mass exceeds the critical value, equilibrium can never be attained, very massive stars probably end catastrophically.

The partial pressure of the ions in the interior of a white dwarf is much smaller than that of the electrons, and if the ions could move freely they would fall into the center of the star, leaving the electrons behind Some charge separation does in fact occur, and gives rise to a radial electric field. It is this electric field that balances the gravitational forces acting on the ions

The chief mechanism of heat flow in white dwarfs is electron conduction, which in all other kinds of stars is negligible compared with radiative transfer. It owes its importance in white dwarfs to the extraordinary mobility of electrons in a degenerate gas. [D L]

Stellar magnetic field

The Sun and many other stars possess magnetic fields similar to that of Earth (*see* GEOMAGNETISM). Since 1945 it has become increasingly evident that gross electromagnetic phenomena are of universal importance in the physics of the stars and of the rarefied gases in the space between them (*see* MAGNETOHYDRODYNAMICS)

Detection of magnetic fields of the Sun and stars and measurement of their intensities in gauss and their polarity as positive or negative or as north or south are accomplished by collecting the light with a telescope, analyzing it with a spectrograph, and studying the Zeeman effect in the spectral lines (*see* ZEEMAN EFFECT). Light-absorbing atoms in the atmosphere of a star spin about the lines of force of the stellar magnetic field so that the spectral lines are split into polarized components. The astronomer can photograph the spectrum and measure the Zeeman splitting, unless the lines are too much broadened by other effects such as the Doppler effect in a rapidly rotating star (*see* STAR).

Earth has a magnetic field of about 0 6 gauss and it was shown in 1953 that the Sun has a general or dipolar field, observable in high heliographic latitudes, some 2 or 3 times stronger than that of Earth. There are much stronger fields in transitory local magnetic regions, including sunspots, in lower latitudes. G E. Hale, in 1908, showed that sunspots have magnetic fields ranging up to 3000 gauss or more.

Observations with the 100-in. and 200-in telescopes of the Mount Wilson and Palomar Observatories resulted in the measurement of magnetic fields in about 90 of the brighter stars that have sharp spectral lines Field intensity ranges up to 5100 gauss in the A-type star 53 Camelopardalis. Stellar fields vary with time, and most of the variations are irregular, showing that intrinsic changes occur in the stellar atmosphere and in the associated magnetic lines of force. Modulation of the apparent field also may result from axial rotation Eight magnetic stars show periodically varying fields of large and somewhat irregular amplitude; all these show reversals of polarity The periods lie between 4 and 10 days.

Magnetic stars show anomalous abundances of the elements, this is attributed to external nuclear reactions resulting from high-energy acceleration of ions in fluctuating magnetic fields of the stellar atmosphere The field of a star controls the disposition of circumstellar ionized gases and so is of importance in cosmogony, in the evolution of stellar systems, and in acceleration of cosmic rays. [H W B]

Bibliography: H W Babcock, A catalog of magnetic stars, *Astrophys J*, suppl 3(30), 1958, H W Babcock, Magnetic fields of the A-type stars, *Astrophys J*, 128·228–258, 1958, T. G. Cowling, *Magnetohydrodynamics*, 1957; S. Fluegge (ed), *Handbuch der Physik*, vol 51, 1958; G P. Kuiper (ed), *Stars and Stellar Systems*, 1960

Stem (botany)

The stem is the organ of vascular plants that usually develops branches and bears leaves and flowers. On woody stems a branch that is the current season's growth from a bud is called a twig. The stems of some species produce adventitious roots *See* ROOT (BOTANY)

GENERAL CHARACTERISTICS

Position. While most stems are erect, aerial structures, some remain underground, others creep over or lie prostrate on the surface of the ground

subtending leaves, and seldom develop immediately after initiation, but remain dormant. If the axillary bud develops into a branch, the latter is gradually organized to resemble the terminal shoot in structure

Accessory or supernumerary buds. In some species, additional buds are produced above or beside the axillary bud. These buds may remain dormant, develop as flower buds, or may grow vegetatively like the axillary buds.

Adventitious buds Adventitious buds are those produced from differentiated (mature) tissues of root, stem, or leaf, or from wound callus In angiosperms they have been observed on hypocotyls (seedling stem below the cotyledons), but rarely on stems

Bud development. The majority of the buds of a plant remain undeveloped for indefinite periods and are known as dormant, latent, or potential buds in contrast to active or developing buds.

In many trees and shrubs, especially those with large, scaly, overwintering terminal or subterminal buds, the whole year's growth is laid down in rudimentary form in the bud during the early summer. The next spring these terminal or upper buds, which are generally more vigorous than those lower down, expand by elongation of internodes Some buds, especially flower buds, may form in the spring of the same season during which they expand In trees, most of the lateral buds remain dormant for the first season, and only a few grow into branches the following spring because of the apical dominance of the terminal bud, which may extend 3–4 ft downward By competing for nutritive or other growth factors, or by forming inhibitory concentrations of growth hormones, the terminal bud may prevent even the formation of lateral buds (see PLANT HORMONES) Conversion of a vegetative terminal bud to a flower bud, or its wounding or decapitation, releases axillary buds (usually the most apical) from inhibition. Ordinarily the most vigorous of the axillary shoots that develop rees- [illegible]

[illegible] appl [illegible] are wounded or pollarded (cut back), originate from adventitious buds

The stems of some plants, as those of most perennial herbs, die down to or beneath the surface of the ground In such plants, lower and older axillary buds are more vigorous than the terminal, and these alone develop into branches the next spring

Stem form and branching. The large and conspicuous stems of trees and shrubs assume various forms

Columnar stems This type of stem is cylindrical, unbranched, and usually bears at its summit one set of large leaves, as palms and bamboos, or no leaves at all, as cacti (see BAMBOO; CACTUS; PALM)

Branching stems. These stems are of two types, excurrent and deliquescent.

The excurrent stem has one main vertical stem or trunk which tapers toward the tip. Typically among the horizontal branches, the longest and oldest are at the base and the shortest and youngest uppermost so that the plant has a conical form, as in pine.

In the deliquescent type of stem, exemplified by elm, the vertical main stem rises for a short distance, then divides into offshoots which branch again and again so that a main axis is no longer evident.

There are three major types of branching: dichotomous, monopodial, and sympodial. Dichotomy is the simplest type and occurs by division of the apical growing point into two equal forks. Dichotomy is rare in angiosperm stems. If the terminal bud of the central axis or a main branch maintains apical dominance and the axillary buds form only lateral and subordinate branches, the branching is called monopodial (Fig. 5). If the

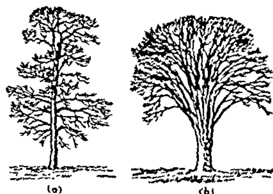


Fig 5 Method of branching in typical trees. (a) Scarlet oak, *Quercus coccinea*, monopodial branching (b) American elm, *Ulmus americana*, sympodial branching (From E W. Sinnott and K. S. Wilson, Botany Principles and Problems, 5th ed , McGraw-Hill, 1955)

apical bud terminates growth as a flower bud, or dies back, or is injured, or otherwise loses dominance, the subsequent growth by the axillary buds is called sympodial branching.

Trees such as pine, spruce, or fir have excurrent form and monopodial branching, whereas horse chestnut, maple, or apple have deliquescent form and monopodial branching. Plants such as elm, lilac, and catalpa have a deliquescent form and sympodial branching

Leafy stems that are indefinite or unlimited in growth are termed long shoots; those definite or limited in growth, characterized by unelongated internodes causing the leaves to appear as though arranged in whorls, are called short shoots Short shoots usually are sympodial systems specialized for reproduction, as the spur shoots of apple. In a sympodial system, the branch appears single, but actually it is composed of a series of lateral branches arranged in lineal order. Reversions to

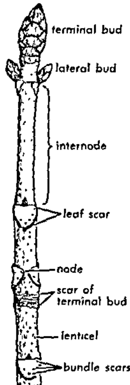


Fig. 3. A woody twig in winter condition (horse chestnut) showing characteristic stem markings and apical dominance. (From E W Sinnott and K. S. Wilson, Botany. Principles and Problems, 5th ed, McGraw-Hill, 1955)

plants, the point of former attachment of leaf is marked on the stem by the leaf scar, where before leaf fall the abscission (separation) layer was formed at the base of the leaf petiole (*see* PLANT GROWTH). Within the leaf scars are small bundle scars left by the broken ends of the vascular (conducting tissue) connections between stem and leaf The shape of the leaf scar is rather characteristic for a given species and may help in its identification. Flower and fruit scars, marking the points of attachment of flowers and fruits, may also be visible, usually just above or to the side of the axillary bud. *See* FLOWER (BOTANY); FRUIT (BOTANY) Whereas nodes are conspicuous in dicotyledons, they are often indistinct among monocotyledons, in which the broad leaf bases may be set in such close succession that no distinction between nodes and internodes is apparent.

Internodes. The stem regions between nodes are called internodes The length attained varies greatly among stems, in different parts of the same stem, and under different growing conditions. In general, the growth of the internode is related to the development of the subtending leaf and each internode grows independently. The order of internode development in dicotyledons is from below upward. A single internode may also elongate first below then above, but often, especially in many monocotyledons, the principal elongation occurs at the base of each internode.

Lenticels. These are small loosely arranged masses of cells in the bark, often slightly raised or ridged (Fig. 4). From the surface they appear as transverse or longitudinal lens-shaped structures Lenticels usually arise beneath stomata. Their intercellular spaces are continuous with those in the interior of the stem, therefore they may be concerned with gas exchange. Their size and shape vary among species and with age. Some few species lack lenticels. *See* PERIDERM.

Bud scales. Buds in perennial plants are protected by a series of modified leaves or bud scales, which in turn are often covered with hairs or wax Special structural arrangements such as enclosure by the growth of the base of the leaf stalk, are found in some plants, especially in the tropics. Buds with protective structures are called covered or protected buds. Herbaceous plants have naked buds, that is, without scales or other protection.

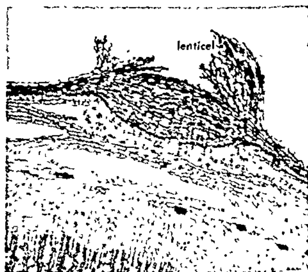


Fig. 4 Portion of a cross section of a young stem of elder, *Sambucus canadensis*, showing a lenticel (From G H Conant, Triarch Products)

Bud scale scars. Narrow ringlike markings in the bark at the limit of each year's growth are called bud scale scars Commonly, these markings are the scars left by the fallen bud scales of the terminal bud of the preceding year The bud scale scars occur in sets because the internodes between the scales do not elongate. The age and yearly growth in length of a branch can be determined by observing the number of sets of these scars and the distances between them.

Bud types. Buds may be classified as terminal, axillary, accessory or supernumerary, and adventitious. Buds are partly developed shoots or undifferentiated masses of meristematic cells. *See* BUD (BOTANY).

Terminal or apical buds These are buds at the tips of branches, often the largest buds on the plant, which usually contribute to the growth in length of the stem bearing them

Axillary or lateral buds Axillary buds are those located on the sides of the stem in the axils of leaves. They are initiated exogenously (in superficial tissues) at nodes, somewhat later than the

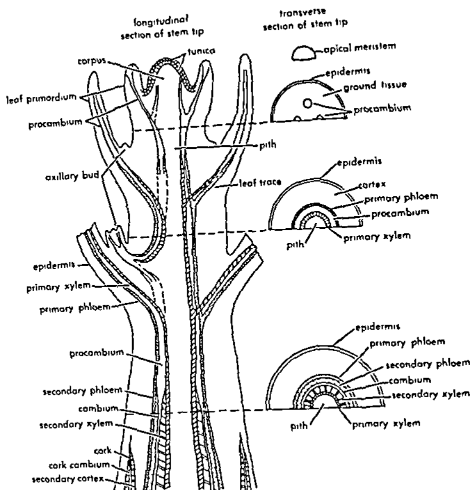


Fig 6 Diagrammatic representation of a woody, dicotyledonous stem in transverse and longitudinal section It illustrates the gradual transition from primary growth (growth in length) to secondary growth (growth in diameter) The primary plant body is derived from the apical meristem, the secondary plant body is produced by the vascular cambium and the cork cambium

cells with stomata, or the diverse epidermal appendages called hairs, or trichomes The walls of the epidermis are cutinized, that is, impregnated with a waxy substance, and the outer walls are covered with a continuous layer of the same substance, the cuticle

Cortex Between the epidermis and the vascular system is the cortex, formed by the outer ground tissue However there is no distinguishable cortex in plants in which the vascular bundles have a scattered arrangement, as in corn The outer layers of the cortex frequently form collenchyma, a mechanical, or strengthening, tissue (*see* COLLENCHYMA) The cortex may also include chlorenchyma (tissue containing chlorophyll) Still other cells may develop as sclereids, fibers, or laticifers Secretory canals may be present. In water plants, the cortical parenchyma may contain large air spaces and is then called aerenchyma The outermost layers of the cortex are sometimes cutinized, like the epidermis and are referred to as the hypodermis. Cortical parenchymal cells contain various inclusions such as starch, tannins, and crystals

Endodermis. The innermost layer of the cortex (or possibly, in the lower vascular plants, the outermost layer of the stele) forms a histochemically distinct and sometimes a morphologically specialized layer, the endodermis If an endodermis is morphologically distinct, it forms a compact layer and the anticlinal walls of its cells bear a suberized and lignified band, or Casparian strip. The endodermis is found in the stems and roots of lower vascular plants, in the rhizomes and roots of seed plants, and in a few herbaceous stems of angiosperms It is not differentiated in the aerial stems of gymnosperms and most angiosperms (*see* PLANT KINGDOM). In young stems of angiosperms a homologous layer contains abundant starch and is called the starch sheath

Pericycle. If a distinct layer or layers of cells are present between the ground tissue endodermis and the vascular tissue, the tissue constitutes the pericycle This tissue is interpreted as part of the ground tissue of the stele. The pericycle may consist of parenchyma or sclerenchyma It may be seen in the stems and roots of lower vascular plants, but is

vegetative long shoots may occur in short shoots.

Among the external stem characters, the arrangement of leaves, the leaf scars, the number and arrangement of axillary buds, the stem form and method of branching, and the bark pattern are all good taxonomic characters (*see* BARK; PLANT TAXONOMY; TREE).

INTERNAL FEATURES OF STEMS

Tissue systems. The stem is composed of the three fundamental tissue systems that are found also in all other plant organs: the dermal (skin) system consisting of epidermis in young stems, of periderm in older stems of many species; the vascular (conducting) system consisting of xylem (water conduction) and phloem (food conduction); the fundamental or ground tissue system consisting of parenchyma and sclerenchyma tissues in which the vascular tissues are embedded (*see* EPIDERMIS, PLANT; PARENCHYMA; PHLOEM; SCLERENCHYMA; XYLEM). The arrangement of the vascular tissues varies in stems of different groups of plants but frequently these tissues form a hollow cylinder enclosing a region of ground tissue called pith and separated from the dermal tissue by another region of ground tissue called cortex (*see* CORTEX, PLANT; PITH; VASCULAR BUNDLES).

Primary and secondary tissues. Part of the growth of the stem results from the activity of the apical meristem located at the tip of the shoot (*see* MERISTEM, APICAL). The derivatives of this meristem are the primary tissues: epidermis, primary vascular tissues, and the ground tissues of the cortex and pith. In many species, especially those having woody stems, secondary tissues are added to the primary. These tissues are derived from the lateral meristems oriented parallel with the sides of the stem: cork cambium (phellogen) that gives rise to the secondary protective tissue, periderm, replacing the epidermis, and vascular cambium, which is inserted between the primary xylem and phloem and forms secondary xylem (wood) and secondary phloem (*see* MERISTEM, LATERAL).

Stele. The vascular tissues and the closely associated ground tissues—pericycle (on the outer boundary of vascular region), interfascicular regions (medullary or pith rays), and frequently also the pith—may be treated as a unit called the stele (*see* PERICYCLE; STELE). The variations in the arrangement of the vascular tissues serve as a basis for distinguishing the stelar types. The word stele means column and thus characterizes the system of vascular and associated ground tissues as a column. This column is enclosed within the cortex which is not part of the stele.

PRIMARY STATE OF STEMS

The leafy stem of the plant is initiated as a shoot primordium in the embryo. This primordium is called the plumule, or epicotyl, and is located above the insertion of the cotyledons (seed leaves) on the hypocotyl. The epicotyl has an apical meristem that gives rise to all subsequent primary parts of the leafy stems. In the primary growth of the stem, nodes and internodes are differentiated and the leaves appear in a characteristic arrangement (phyllotaxis). The cells derived from the apical meristem compose the meristematic precursors of the primary tissues of the stem and are called primary meristems. These meristems differentiate into the primary tissue systems.

Shoot apex or apical meristem. This so-called growing point consists of the meristematic initials and their derivatives. The organization of the apical meristem varies in different groups of plants. In the angiosperms the initials are usually arranged in two or more layers (tiers), frequently three; in most gymnosperms in one layer; and in the ferns and still lower vascular plants the shoot tip may bear a single initial. The organization in the angiosperms is referred to as that of tunica and corpus. The outermost tier or tiers of initials and their immediate derivatives constitute the tunica, the innermost tier and its immediate derivations compose the corpus. The tunica is a mantlelike region enclosing the corpus or core. The tunica increases only in surface by anticlinal (at right angles to the surface) divisions; the corpus shows volume growth, for its cells divide in various planes. All the initials and their first derivatives together are often designated as a promeristem. Below this promeristem are partially differentiated tissues, the primary meristems, named from the periphery inward: the protoderm, the outer ground meristem, the procambium, or provascular tissue, and the inner ground meristem. As stated previously, the primary meristems differentiate into primary tissue systems, the protoderm into the epidermis, the ground meristem into tissues of cortex and pith, and the procambium into primary vascular tissues.

Differentiation. The complex of physiological and morphological changes that occur in a cell as it develops from a meristematic to a mature state is called differentiation. Common evidences of differentiation of cells are vacuolation (water uptake), accumulation of ergastic substances (metabolic products), formation of plastids, increase in thickness of the cell wall and its impregnation with new wall substances. In highly specialized cells the nucleus may break down (sieve elements in the phloem) or even the entire protoplasm may be lost (water-conducting cells in the xylem). *See* CELL (BIOLOGICAL).

Primary tissue systems. As seen in the stem as a whole, the primary tissues form integrated systems and are referred to as the primary tissue systems: the epidermis (or epidermal system), the ground tissue system, and the primary vascular system.

Epidermis. The epidermal system differentiates from the protoderm. The epidermis is usually composed of a single layer of cells. Occasionally, the protoderm divides periclinally (parallel to the surface) and produces a multiple epidermis. Epidermal cells may be variously modified into guard

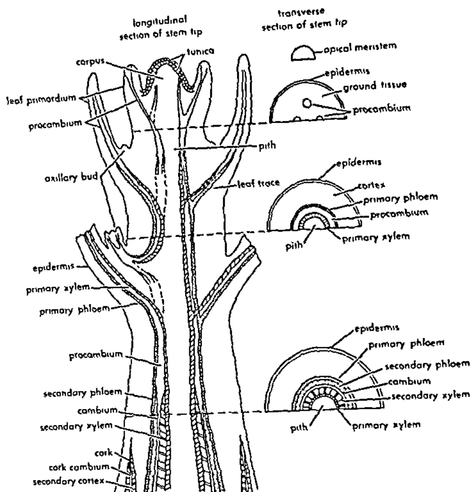


Fig 6 Diagrammatic representation of a woody, dicotyledonous stem in transverse and longitudinal section It illustrates the gradual transition from primary growth (growth in length) to secondary growth (growth in diameter). The primary plant body is derived from the apical meristem, the secondary plant body is produced by the vascular cambium and the cork cambium.

cells with stomata, or the diverse epidermal appendages called hairs, or trichomes The walls of the epidermis are cutinized that is, impregnated with a waxy substance and the outer walls are covered with a continuous layer of the same substance, the cuticle

Cortex Between the epidermis and the vascular system is the cortex, formed by the outer ground tissue However there is no distinguishable cortex in plants in which the vascular bundles have a scattered arrangement, as in corn The outer layers of the cortex frequently form collenchyma, a mechanical, or strengthening, tissue (see COLLENCHYMA) The cortex may also include chlorenchyma (tissue containing chlorophyll) Still other cells may develop as sclereids, fibers or laticifers Secretory canals may be present In water plants the cortical parenchyma may contain large air spaces and is then called aerenchyma The outermost layers of the cortex are sometimes cutinized, like the epidermis, and are referred to as the hypodermis Cortical parenchymal cells contain various inclusions such as starch, tannins, and crystals

Endodermis The innermost layer of the cortex (or possibly, in the lower vascular plants, the outermost layer of the stele) forms a histochemically distinct and sometimes a morphologically specialized layer, the endodermis If an endodermis is morphologically distinct, it forms a compact layer and the anticlinal walls of its cells bear a suberized and lignified band, or Casparian strip The endodermis is found in the stems and roots of lower vascular plants, in the rhizomes and roots of seed plants, and in a few herbaceous stems of angiosperms It is not differentiated in the aerial stems of gymnosperms and most angiosperms (see PLANT KINGDOM) In young stems of angiosperms a homologous layer contains abundant starch and is called the starch sheath

Pericycle If a distinct layer or layers of cells are present between the ground-tissue endodermis and the vascular tissue, the tissue constitutes the pericycle This tissue is interpreted as part of the ground tissue of the stele. The pericycle may consist of parenchyma or sclerenchyma. It may be seen in the stems and roots of lower vascular plants, but is

usually absent in the stems of gymnosperms and angiosperms The fibers located on the periphery of the vascular systems in many dicotyledons are often called pericycle fibers. In many species, however, these fibers originate in the phloem.

Pith. The ground tissue enclosed by the vascular tissues is the medulla or pith. In stems having no cortex, the pith also is not distinguishable. The vascular bundles may be scattered throughout the central ground tissue, as in some monocotyledons; or there may be strands of internal phloem or complete vascular bundles located within the pith, as in some dicotyledons. The pith may consist of parenchyma or may become sclerotic (hard-walled), especially at the nodes. The outer zone of the pith frequently consists of small rather persistent cells with more or less sclerified walls Many herbaceous plants have hollow stems (internodes) because the growth of the pith fails to keep pace with stem expansion Some woody plants, as the walnut and *Liriodendron*, have a diaphragmed pith, that is, pith with thick-walled parenchyma cells and sclereids arranged in disks, or diaphragms, of firm tissue.

Primary vascular system. In the ferns and the seed plants the primary vascular system, derived from the procambium, is composed of anastomosing (interconnected) vascular bundles At each node some of the vascular bundles diverge into the leaves and axillary branches A bundle in the stem which leads directly into a leaf is called a leaf trace; one leading into a branch is a branch trace One or more leaf traces may be associated with each leaf Where the leaf trace diverges from the stem into a leaf at a node, the vascular cylinder commonly has a more or less circumscribed region of parenchyma This region is called a leaf gap or lacuna. The lateral branches of most dicotyledons and gymnosperms have two branch traces Some species have one trace per branch, others more than two. In monocotyledons, for example, the axillary shoot connection consists of many vascular strands At the node, the branch traces may have a single branch gap which, moreover, is commonly confluent with the leaf gap.

Separate parts of the primary vascular system of the stem constitute vascular strands or vascular bundles. The phloem and xylem within a vascular bundle may be arranged in various ways. The most common type in gymnosperms and angiosperms is the collateral bundle in which the phloem occurs only on one side of the xylem If phloem occurs on both sides of the bundle, as in the Cucurbitaceae, the bundle is bicollateral. If one kind of vascular tissue completely surrounds the other, the bundles are concentric. The concentric bundle is amphicribral if the phloem surrounds the xylem, as in ferns, or amphivasal if the xylem surrounds the phloem, as in some monocotyledons and dicotyledons.

The primary vascular system and associated fundamental tissue is termed the stele The simplest type of stele, and perhaps also the most primitive, is the protostele, in which a solid central core of xylem is surrounded by phloem, as found generally in the roots, in the stems of lower vascular plants such as the fern *Gleichenia*, and the stems of some water plants (Fig. 7). In the plant group Lycopsida, the lycopods or "ground pines" have a modified protostele with alternating bands of xylem and phloem, and the "club mosses" (*Selaginella*) have separate protosteles supported in air spaces by radially elongated endodermal cells designated trabeculae. In the stems of vascular plants located higher than Lycopsida on the evolutionary scale, the center of the stele is occupied by a pith and the vascular system has a tubular structure. The stele, therefore, is called a siphonostele (tubular stele). If the phloem occurs only on the outer side of the xylem of a siphonostele (gymnosperms and angiosperms), the stele is called an ectophloic siphonostele. If the phloem differentiates also on the inner side of the xylem (in such ferns as *Adiantum*, *Dicksonia*, and *Marsilia*) the stele is designated an amphiphloic siphonostele or a solenostele In the ferns, in which the stem is short and the leaves are set close together, the leaf gaps overlap and the tubular stele is dissected into a network of distinct strands Such a stele is called a dictyostele as in *Polypodium* and *Dryopteris*. Another modification of the siphonostele is the eustele, in which the vascular system consists of collateral or bicollateral strands and both interfascicular regions and leaf gaps are present Eustele is characteristic of gymnosperms and angiosperms. The most complex stele, the atactostele, has a dispersed arrangement of strands, as in many monocotyledons

Primary development. The origin of the stem from the apical meristem is closely associated with the development of leaves At initiation, the leaf primordium is merely a small protuberance on the flank of the shoot tip and cannot be delimited from the stem part to which it is attached Only later, as growth continues, does the primordium become an organ distinguishable externally and internally from the stem The stem itself shows at first no division into nodes and internodes It appears like a complex of superimposed disks (future nodes), each bearing one or more leaf primordia depending on leaf arrangement. Later during elongation, the disklike zones become separated from one another by cell division and enlargement between the levels at which the leaf primordia are inserted Thus the internodal elongation is initiated and the stem becomes differentiated into nodes and internodes The elongation of internodes is the phenomenon that brings about the characteristic rapid extension of new shoots during the spring growth of trees and shrubs

Internally the differentiation of stem tissues is more or less closely correlated with the differentiation of leaves In lower vascular plants with protosteles in their stems, as the lycopods the vascular system differentiates beneath the apical meristem as a column of provascular tissue (procambium), and the leaf traces, which connect the small leaves with this column, differentiate later through the cor-

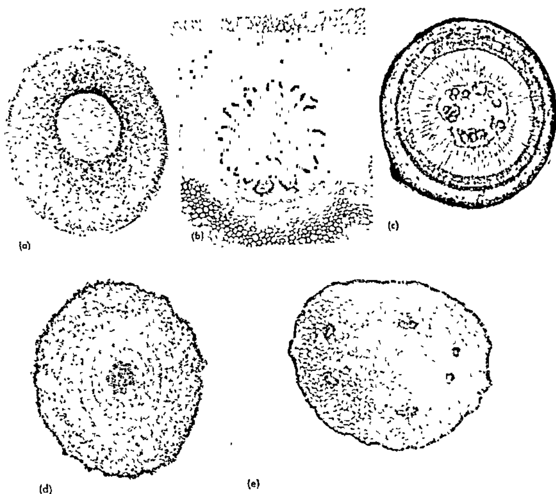


Fig 7 Types of steles as seen in photomicrographs of cross sections of stems (a) Protostele (*Gleichenia*) (b) Radial (modified) protostele (*Lycopodium*). (c) Ectophloic siphonostele (*Tilia*). (d) Amphiphloic siphonostele (*Dicksonia*) (e) Dictyostele (*Polypodium*).

tex In its differentiation the provascular tissue becomes distinct from the surrounding ground tissue of the cortex because the cells of the latter show an increase in vacuolation and an overall enlargement whereas the cells of the provascular tissue retain their dense cytoplasm and increase in length but remain narrow (see CYTOPLASM) In the ferns and the seed plants in which the leaves are more prominent parts of the shoot than in the lycopods and the vascular system of the stem consists of a cylinder of strands (siphonostele structure) connected with the leaves at regular intervals (at the nodes) the vascular system is initiated in relation to the leaves (see ANGIOSPERMAE, FILICALES, LYCOPODIALES) Vacuolation of the differentiating cortex and pith delimits the future vascular region as a still highly meristematic cylinder of cells Then within this cylinder, procambial strands differentiate beneath the emerging leaf primordia chiefly by longitudinal cell division and cell elongation, that is the leaf traces are initiated. The cells between the bundles that do not become procambial cells differentiate as parenchyma cells of the interfascicular regions and the leaf gaps. If followed downward in the stem, the young leaf traces will be found joined with one another and with older traces Thus, the procambial strands form an interconnected system similar to that of the mature primary vascular system, except that all its parts are still short The procambial system elongates in coordination with the other tissues of the internodes by further division and elongation of cells Eventually, primary phloem and primary xylem differentiate in the procambial strands, the phloem in continuity with that of the older parts of the stem (acropetal differentiation, that is, from the base of the stem toward its apex), the xylem first appearing at the bases of leaves then differentiating in two directions downward (basipetally) toward a connection with the xylem of older stem regions, and upward in the leaves.

Differentiation of lateral buds. The lateral or axillary buds arise more or less close to the apical meristem of the terminal bud depending on the type of branching characteristic of the plant. The buds are initiated by cell division in the stem near the axils of subtending leaves The divisions occur in such a manner that eventually an apical meristem

of the bud is organized. The bud meristem may develop immediately into a side shoot by producing leaves and increments of stem, or it may interrupt growth after variable degrees of initial development. The lateral bud may also develop into an inflorescence or flower by appropriate changes in growth pattern as compared with that of a vegetative shoot. The bud and parent stem establish a vascular connection through the bud traces, the differentiation of which is more or less similar to that of the leaf traces of the same plant.

SECONDARY STATE OF STEMS

The stems of gymnosperms, most dicotyledons, and some monocotyledons show an increase in thickness by secondary growth. A lateral meristem, the vascular cambium, produces secondary vascular tissues which constitute the secondary body. Also, a cork cambium, or phellogen, produces a secondary protective tissue called periderm, a tissue which replaces the primary protective tissue, the epidermis.

The vascular cambium originates from the procambial and certain parenchyma cells of the primary body. Procambial cells that remain meristematic after the stem completes its elongation and the primary vascular tissues become mature, proceed to divide periclinally (parallel with the surface). These cells are designated as cambial initials. In addition, periclinal divisions are initiated by parenchymal cells between the vascular bundles; these cells also are cambial initials. Collectively all these initials constitute the vascular cambium. The part of cambium originating within the vascular bundles is the fascicular cambium; that arising in the parenchyma between vascular bundles, the interfascicular cambium (Fig. 8).

The arrangement of the secondary vascular tissues produced by the cambium varies in different plants. This tissue may appear as (1) a continuous cylinder, (2) individual strands with secondary activity limited to the bundles (no interfascicular cambium is formed), (3) individual strands, with secondary vascular tissues produced in the fascicular regions and secondary ray parenchyma in the interfascicular regions, or (4) anomalous (atypical) arrangements characterized by uneven growth of xylem or phloem or by formation of successive cambia, one farther outward than the preceding.

The cork cambium or phellogen originates in the epidermis or in the cortex from parenchyma or collenchyma cells. It produces phellem or cork outwardly and phelloderm or secondary cortex inwardly.

Woody dicotyledons. The cambium is a continuous cylinder. The fascicular cambium produces secondary xylem and phloem, the interfascicular cambium forms either the same kind of tissue as the fascicular cambium or only vascular ray tissue.

As seen in cross sections, the secondary xylem, or wood, is formed in distinct concentric rings called annual, or growth, rings (Fig. 9). Each annual ring is composed of an inner portion called the spring, or early, wood, in which the vessels and

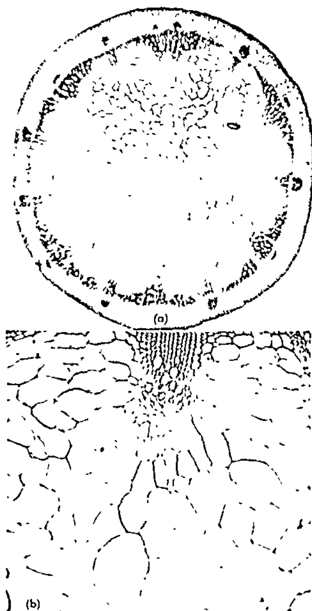


Fig. 8 Secondary tissue in the stem of sunflower (*Helianthus annuus*). (a) Transverse section of the entire stem showing the thicker walled, dark, secondary xylem cells as they are added to the primary, more in the fascicular than in the interfascicular areas. (b) Enlarged view of a portion of a. (From J. B. Hill, L. O. Overholts, and H. W. Popp, Botany, A Textbook for Colleges, 2d ed., McGraw-Hill, 1950)

tracheids are relatively large, and an outer portion called the summer, or late, wood, in which the cells have much smaller diameters and thicker walls than those of the early wood. Usually only the outermost growth rings function in conduction of water. The outer, light colored, relatively soft, functioning part of the wood is called sap wood. The nonconducting older (early formed) wood, called heart wood, is often filled with gums, resins, tannins, and mineral salts, and is dark in color. Growth rings in oblique or horizontal branches become asymmetrical, the rings being wider in the upper half (in most angiosperms) or in the lower half (in conifers). The wood with the wider rings has a different histologic and chemical structure from that with the

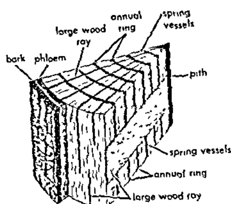


Fig 9 A diagram of a segment of an oak log showing in three dimensions the grain, that is, the pattern formed by the annual rings and the wood rays. At the right, the block has been cut radially, and above, transversely At the left (the surface of the log), a portion of the phloem and bark has been removed, showing a tangential view of the wood underneath. (From E W Sinnott and K S Wilson, Botany, Principles and Problems, 5th ed, McGraw-Hill, 1955)

narrower and is called tension wood in the angiosperms and compression wood in the conifers.

The wood varies in structure chiefly depending on (1) the relative amounts of vessels tracheids, and wood fibers, (2) the distribution of wood parenchyma, and (3) the presence and character of vascular rays

Two patterns of vessel arrangement are recognized If the early wood contains wide vessels and the late wood narrow vessels, so that each growth ring begins with a zone of large vessels, the wood is called ring-porous (Fig 10) If vessels of more or less uniform widths are formed throughout the year, the wood is called diffuse-porous

Wood texture, coarse or fine refers to the size and number of the xylem cells, especially vessels.

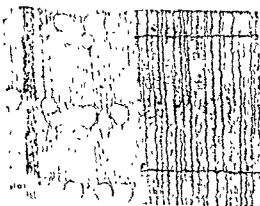


Fig 10 Transverse sections, much magnified, of two kinds of dicotyledonous wood (a) White oak, a ring-porous wood with very large spring vessels, and narrow (uniseriate) and wide (multiseriate) rays. (b) Yellow birch, a diffuse-porous wood with narrow (mostly uniseriate) rays (U.S. Forest Products ...

The term grain refers to the arrangement of the cells, especially fibers, making up the wood. A wood is hard or soft depending upon the number of lignified fibers

Among the many wood products, plywood and veneer are particularly important. In making plywood, the wood is softened and the log rotated against a heavy blade to produce a continuous sheet of very thin wood which is cut to suitable lengths, pieces alternately stacked with grain at right angles are bonded together in varying odd numbers of layers, pressed, and dried Plywood, or other wood bases, may be covered by thin sheets of fine-grained wood or veneer. Characteristic and interesting grain patterns for fine woods or veneers are obtained by cutting wood in transverse, radial, and tangential section (Fig 11). Veneer wood with the

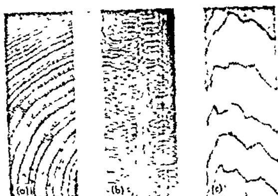


Fig 11. Appearance of silver maple (Acer saccharinum) wood in different sections. The dark and light colors and the grain irregularities result from cutting through the fibers and vessels of the annual rings in different planes In addition, the rays, particularly in radial section, give the wood a pleasing and interesting appearance. (a) Transverse section (b) Radial section (c) Tangential section (From J. B Hill, L. O Overholts, and H W. Popp, Botany, A Textbook for Colleges, 2d ed, McGraw-Hill, 1950)

most interesting pattern or grain comes from the base of a tree, from burls or from crotch wood where the vascular arrangements may be wavy or spiral Some of the choice curly grains such as curly maple result from the development of adventitious buds. Injury to the cambium may produce a similar ...

... cylinder of secondary xylem and phloem. Commercial gymnosperm wood is usually derived from conifers Conifer wood is composed of tracheids, fiber-tracheids, and uniseriate wood rays but no vessels (Fig 12) Because of the absence of vessels the conifer wood is more homogeneous in sections than the angiosperm wood (Fig 12)

As ...

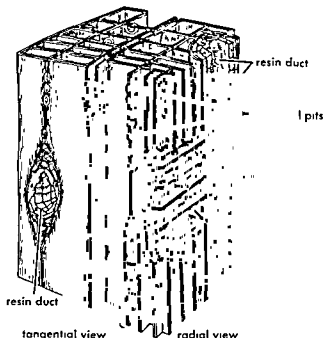


Fig. 12. A stereogram of a small block of white-pine wood as seen in three dimensions. It illustrates the relationships of the vertical tracheids and the ray elements. (*From J B Hill, L. O Overholts, and H. W Popp, Botany, A Textbook for Colleges, 2d ed, McGraw-Hill, 1950*)

forming knots Knots weaken the wood and are not desirable unless for ornamental paneling

The secondary growth of other gymnosperms is similar to that of the conifers The cycads are interesting because growth rings are added at intervals of 2–20 years and are difficult to distinguish except in *Dioon*. Many of the cycads form only one ring of wood; but others, such as *Cycas*, *Dioon*, *Macrozamia*, and *Encephalartos*, produce several rings.

Monocotyledons. Some monocotyledons, as the bamboos and palms, may form rather woody trunks by primary thickening The thickening growth may occur rather generally throughout the stem or there may be a primary thickening meristem, a cambiumlike zone of cells originating from the ground tissue beneath young leaf bases. It increases the width of the stem directly beneath the apex Additional division of cells throughout the stem and the enlargement of cells complete the growth in thickness. In certain monocotyledons, such as *Aloe, Yucca, Agave, Dracaena*, secondary growth also occurs. However, the cambium does not originate in the bundles, which are scattered in the ground tissue. It appears in the ground parenchyma outside the vascular bundles and by tangential divisions forms parenchyma on the outside and both vascular strands and parenchyma on the inside. The basic structure of the primary and secondary bodies is the same—in both, discrete vascular strands are embedded in ground tissue, usually sclerified (hardened) parenchyma.

Effect of secondary growth. The secondary xylem covers the primary xylem and pith, usually without changing them, whereas the primary phloem and cortex are pushed outward and are

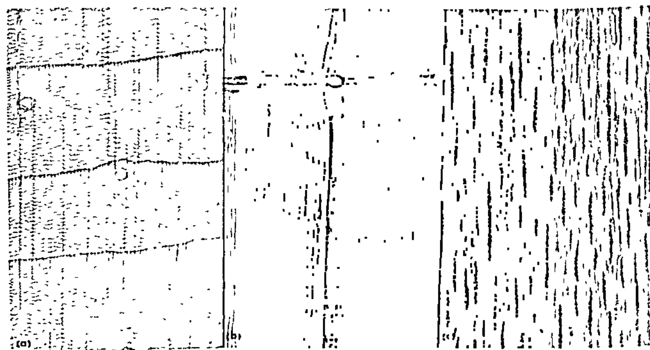


Fig 13. Enlarged photomicrographs of pine wood (a) Transverse section. The wood cells are here cut across at right angles to their length. *Note the thin-walled cells (tracheids) in the springwood and the thick-walled ones in the summerwood (fiber-tracheids)* The uniseriate wood rays run at right angles to the annual rings. The large openings are resin canals (b) Radial section (c) Tangential section In the last two sections the tracheids are cut lengthwise In the *radial section the rays run horizontally, across the* tracheids In the tangential section, the rays appear as lens-shaped clusters of cells (*U S Forest Products Laboratory*)

more or less compressed by the centrifugal growth of the secondary vascular tissues The epidermis and cortex may keep pace with secondary vascular growth, or they may be ruptured and replaced by a periderm. Cambium develops in the parenchyma of the leaf gap and its derivatives close the leaf gaps and break and bury the leaf traces

STEM TYPES

The stems of vascular plants, or Tracheophyta, show an endless variety of form and structure. In this article only a few are considered as type examples See TRACHEOPHYTA

Gymnosperm stems. This group of Tracheo-

annual or perennial herbs are recognized in this group See GYMNOSPERMAE.

Coniferales The gymnosperms of this order are much branched, small-leaved, woody tree forms, except some shrubby junipers. The stem has typically an ectophloic siphonostele Most conifers produce a conspicuous amount of secondary tissues which form a solid continuous cylinder (Fig. 14). Resin ducts occur in the cortex and the vascular tissues, except in the yews (see SECRETORY STRUCTURES, PLANT) The wood is dense and massive, composed of tracheids, fiber-tracheids, and uniseriate wood rays See WOOD (ANATOMY AND IDENTIFICATION).

Cycadales The stem types range from tuberous and partly or wholly subterranean to relatively tall, aerial stems. The stem is usually unbranched and bears a crown of large leaves The cortex and pith are large, loose, and parenchymatous The vascular cylinder is narrow, has broad medullary rays and a loose textured wood. The stele is a siphonostele The primary xylem usually contains scalariform tracheids, the secondary xylem tracheids with bordered pits. Leaf traces girdle the stem, that is, they have a nearly horizontal course in the cortex. The stem is rigid mainly because of persistent leaf bases.

Ginkgoales. The only living representative, *Gingko biloba*, is a large profusely branched tree with both long shoots (normal vegetative) and short shoots (restricted vegetative or reproductive). Secondary growth is vigorous, and the pith and cortex are relatively small.

Gnetales. The three genera included in this order are highly specialized *Ephedra* is usually a much branched shrub. *Gnetum* species are mostly lianas, and in *Welwitschia* most of the stem is buried in the ground A feature which distinguishes these plants from other gymnosperms is the presence of vessels among the xylem elements.

Angiosperms. The angiosperms (flowering plants) include two main groups—the dicotyledons

Although the various stem types intergrade in their characters, some approximate generalizations regarding their differences may be made (see list).

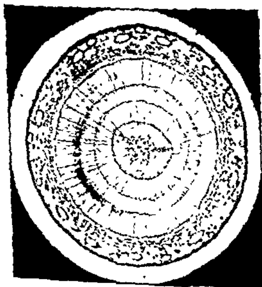


Fig 14 Cross section of the stem of pine, 4 years old, showing annual xylem rings (1–4), resin ducts, and other tissues. (From J B Hill, L. O Overholts, and H W Popp, Botany, A Textbook for Colleges, 2d ed., McGraw-Hill, 1950)

COMPARISON OF STEMS

Dicotyledons	*Monocotyledons*
1. Stems woody, herbaceous, vines	1. Stems herbaceous, few woody
2. Secondary growth common, phloem, vascular cambium, and xylem in concentric layers, separate vascular bundles common in the primary body, uncommon in secondary	2 Some woody types with secondary growth; vascular cambium not between xylem and phloem; in the primary body vascular bundles often scattered through ground tissue, sometimes in two circles; in the secondary body also separate vascular bundles
3. Distinct pith and cortex	3. No definable cortex and pith
4. Endodermis and pericycle usually absent; sometimes pericyclic fibers on outer limit of vascular cylinder, usually phloem fibers	4 Endodermis and pericycle usually absent; vascular bundles with sheaths of sclerenchyma
5 Number of leaf traces to each leaf is variable, rarely large	5. Number of leaf traces to each leaf is large; sheathing leaf bases enclose the stem
6 Periderm frequently present	6. Periderm not common

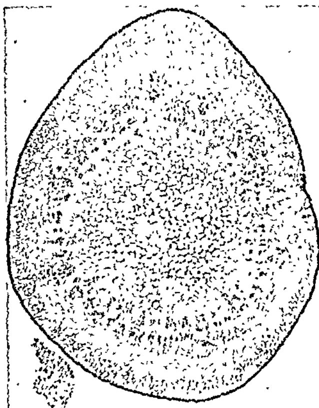


Fig. 15. Woody dicotyledon One-year-old woody stem of *Liriodendron*, showing the characteristics of an ectophloic siphonostele. The epidermis has not yet been replaced by periderm.

Woody dicotyledons. The primary vascular cylinder in the woody dicotyledons is an ectophloic siphonostele, and the pith and cortex are well defined. *Liriodendron* illustrates this type of structure (Fig. 15). The stem, from the inside out, consists of (1) a central pith of large, loose, parenchyma cells; (2) a vascular system containing closely arranged vascular bundles, with the leaf traces often recognizable as rather large bundles protruding into the cortex; (3) a cortex; and (4) an epidermis. When secondary growth occurs a continuous vascular cambium and continuous layers of secondary xylem and phloem are formed The epidermis is eventually replaced by a periderm There is no endodermis and no pericycle Primary phloem fibers form the outer limit of the vascular region (*see* ENDODERMIS).

Woody vines or lianas. Vines are generally characterized by a primary and secondary vascular system dissected into strands by broad rays (Fig 16*a*). Also, they often have a ringbark which results from a concentric arrangement of the successively formed periderms. Lenticels may be absent in ringbark.

Tropical lianas (climbing woody vines) often have a more or less unusual structure. There may be, for example, a highly irregular development of secondary xylem and phloem, so that the outline of the xylem is uneven (Fig. 16*b*). A still more striking structure results when the leaf traces develop complete cylinders of cambium, xylem, and phloem at levels where they have diverged from the central cylinder. The stem appears as though composed of several central cylinders (Fig. 16*c*).

Herbaceous dicotyledons. These forms differ from the arborescent, or woody, dicotyledons chiefly in having a smaller amount of cambial activity or none They are often similar to 1-year-old woody stems. The primary vascular cylinder forms an ectophloic siphonostele with the vascular strands often more widely separated by interfascicular regions than those in the arborescent species (Fig. 17). Vascular cambium may be absent, as in *Ranunculus*, may be present, but form only parenchyma and sclerenchyma in the interfascicular regions, as in *Medicago*; or it may form a continuous vascular cylinder, as in *Vinca* or *Digitalis*

Herbaceous monocotyledons. In these plants the vascular system is composed of widely spaced strands arranged in one of four ways. First, as in most grasses, the vascular bundles are arranged in two circles, with the outer smaller bundles embedded in a continuous sheath of sclerenchyma close to the epidermis (Fig. 18). The vascular bundles are collateral, each enclosed in a sheath of sclerenchyma. The pith may break down in the internodes but not in the nodes. Transverse bundles

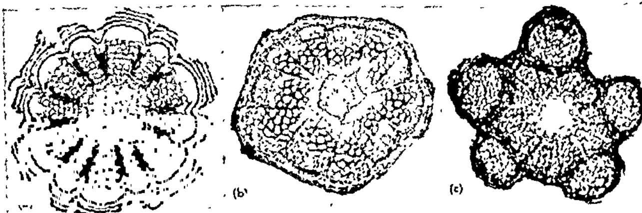


Fig 16. Transverse sections of woody vines. (a) *Clematis* (b) *S. mexicana*. Mature stem showing five well-developed cortical "steles." (c) *Serjania subdentata* Mature stem showing irregular development of secondary xylem and phloem.

in the nodal region interconnect the leaf traces

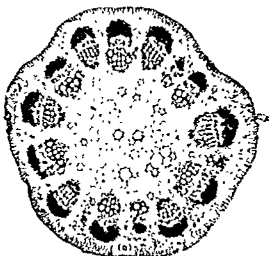
Second, in a few monocotyledons, for example *Clintonia*, a single, concentrically arranged series of bundles may occur Each bundle has its own complete endodermis

Third, in a few monocotyledons the bundles are grouped in the center of the stem, as in the rhizome of *Acorus calamus* The bundles are amphivasal, the phloem being surrounded by xylem.

Fourth and most commonly, vascular bundles are numerous and scattered in a ground tissue, as in corn or bamboo, with no pith or cortical regions being evident. The complex arrangement of the vascular bundles is related to the large number and variation in size of leaf traces There may be a sclerenchyma cylinder about each bundle, or it may be lacking, but subepidermal parenchyma is strongly sclerified Endodermis and pericycle are absent.

Woody monocotyledons Some monocotyledons, as the bamboos and palms, may form rather woody trunks by thickening of cell walls with age until nearly the whole ground tissue is heavily sclerified. Despite their large size these trunks are composed of primary tissues only There are, however, woody monocotyledons with secondary growth, but the secondary tissues resemble the primary in being composed of vascular bundles embedded in ground tissue, usually sclerified parenchyma

Specialized erect stems. Two kinds of erect stems have special names. Culm is a name applied to hollow, but solid-jointed, stems of grasses, woody and herbaceous Caudex means the axis of a plant, consisting of root and stem The term is sometimes applied to short, enduring stems or stocks which throw up new stalks each year from



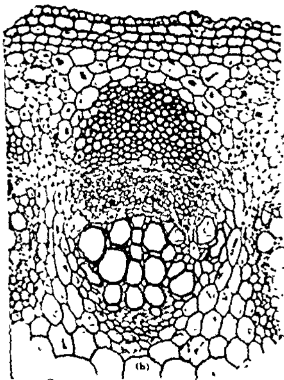


Fig 17 The primary tissues of the stem of a typical herbaceous dicotyledonous plant, sunflower, *Helianthus annuus* (a) Transverse section of entire stem showing arrangement of vascular tissues in one circle (b) Enlarged portion of a showing details of structure, including one vascular bundle (From J B Hill, L O. Overholts, and H W. Popp, *Botany, A Textbook for Colleges*, 2d ed., McGraw-Hill, 1950)

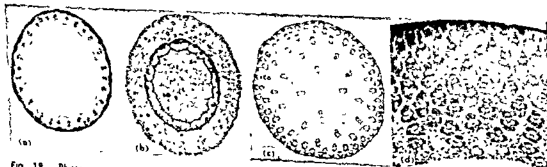


Fig 18 Photomicrographs of transverse sections of monocotyledonous stems showing various arrangements of vascular bundles. (a) Wheat (grass) (b) *Clintonia*. (c) Corn (d) Bamboo (From J. B Hill, L O. Overholts, and H. W Popp, *Botany, A Textbook for Colleges*, 2d ed., McGraw-Hill, 1950)

persistent buds at ground level. Also, the name is used to describe the trunks of palms and tree ferns which consist of persistent leaf bases forming a column.

Perennial stems. Perennials which have woody stems are generally called trees or shrubs. Shrubs are woody perennials with several main stems branched from or near the ground and generally *not more than 20–25 ft high. Trees are woody* perennials with a single main trunk or axis which rises some distance above the ground before branching. Trees generally exceed shrubs in height The two words are convenient but not exact terms because trees and shrubs intergrade in height and form. Herbaceous plants also may be perennial. Their stems commonly die down to the ground each year.

Modified aerial stems. Stems may be variously modified for photosynthesis, storage, propagation, support, and protection.

Cladode. This term is applied to a branch with a single internode which is flattened and serves as a leaf, as in asparagus.

Phylloclade. A flattened leaflike shoot that replaces a leaf as a photosynthetic organ, as in the *Opuntia* cactus is called a phylloclade.

Thorn. This hard, pointed projection may be a modified branch, or leaf, or stipule. Thorns may be unbranched, as in the osage orange, or branched, as in the honey locust. A lateral branch may form leaves and flowers, then terminate growth by producing a thorn at the apex.

Creeping or prostrate stems. Stems that trail along the surface of the ground and may take root at the nodes are called creeping or prostrate.

Runner. A runner, as in the strawberry, is a horizontally growing, sympodial stem system; that is, it appears single but actually is composed of a series of lateral branches arranged in lineal order. In this system the stem forms adventitious roots near the tip, then gives rise to a rosette of leaves from a bud. A new runner emerges from the axil of a reduced leaf at the base of this rosette.

Stolon. The term stolon is used for creeping stems, shorter than runners. They also root adventitiously, as in raspberry or currant. The term has no precise meaning.

Climbing stems or vines. These are long, slender stems which usually climb by special devices.

Rambler. The ramblers rest on the tops of other plants and some, as certain roses, have spines or prickles which help them to adhere to their support.

Root climbers. English ivy and poison ivy have stems which climb by means of adventitious roots.

Tendril climbers. These plants climb by means of modified leaf tendrils, as in the garden pea, or by stem tendrils, as in the grape.

Twiners In the twiners, the whole stem winds about its support, as pole beans or many tropical lianas

Underground stems. A number of stems grow underground and are often mistaken for roots. The principal kinds of underground stems are rhizomes or rootstocks, tubers, corms, bulbs, and rhizomorphic droppers.

Rhizomes are usually quite varied, plagiotropic (growing horizontally), perennial, underground stems found in a vast number of plants. Their cauline (stem) nature is evident by well-defined nodes and internodes. Roots, scale leaves, and axillary buds form at the nodes. Some buds form leafy, upright shoots while others form new, sympodially branched underground shoots. Among different species, rhizomes may be (1) thin, tough, and rapidly growing; (2) fleshy though rapidly growing; and (3) short and fleshy, serving more for storage than for spreading. The anatomy of rhizomes, except for some dorsiventrality, is much like that of the aerial stem of the same species. The other types of underground stems, such as tubers, corms, and bulbs, are essentially modified rhizomes. [J.E.CU.]

Bibliography: See PLANT ANATOMY.

Stem cuttings

Young shoots or sections of older stems used in the practice of asexually reproducing desirable plants by inducing such shoots or stem pieces to develop roots Herbaceous stem cuttings usually root readily in water, moist sand, peat, or vermiculite, but many woody plants require special treatment. In woody evergreens, rooting is more successful in winter than in summer, when stem pieces are taken from young rather than from old plants, and when the cuttings are treated with auxin, usually indolebutyric or indoleacetic acid (100–200 milligrams per liter for 24 hours), or a commercial rooting hormone (*see* PLANT GROWTH, PLANT HORMONES). In deciduous woody plants, the most important factor is age (young shoot with few buds and some leaf area), whether or not auxin is used. For most plants requiring auxin for rooting, aqueous solutions (25–40 milligrams per liter) applied to the base are most effective For plants that are difficult to root, information may be obtained from tables of different auxins showing the concentrations recommended for individual plant species *See* REPRO-



Cuttings of holly. (a) Controls treated at base with talcum powder only and showing no roots (b) Cuttings treated with powder containing indolebutyric acid and naphthaleneacetic acid and rooting vigorously (*From Boyce Thompson Institute for Plant Research in E. W. Sinnott and K. S. Wilson, Botany, Principles and Problems, 5th ed., McGraw-Hill, 1955*)

DUCTION, PLANT; ROOT (BOTANY); STEM (BOTANY) [J.E.GU.]

Bibliography: K. V. Thimann and J. B. Rogers, *The Use of Auxins in the Rooting of Woody Cuttings*, Harvard Univ. Maria Moors Cabot Foundation for Botanical Research, Publ. no. 1, 1947

Stenolaemata

An order of ectoproct Bryozoa proposed by F. Borg in 1926 to include orders Cyclostomata and Trepostomata. The term Stenolaemata is objectionable to bryozoologists both on etymological grounds and because it includes the Treptostomata. E. Marcus, in 1938, suggested Stenostomata as a more suitable name. *See* BRYOZOA. [M.D.RO.]

Stenurida

An extinct order of Ophiuroidea, embracing Paleozoic forms in which the ambulacral groove remained open and the ambulacral plates remained as discrete elements that were not fused together in pairs to form vertebrae. These characters indicate that the Stenurida were intermediate between asteroids and existing ophiuroids. *See* OEGOPHIURIDA; OPHIUROIDEA. [H.B.F.]

Steppe

A regional name, of the great semiarid grassland plains of Eurasia, that has become a descriptive designation of short grass vegetation associated with semiarid climate all over the world. The vegetation is typically composed of a relatively continuous cover of short grass, 1 ft or less in height, occurring in semiarid parts of the mid latitudes. The rainfall fluctuates enormously from year to year but is sufficient in most years to produce a shallow zone of soil moisture during the warm season. On their wetter margins, steppes merge into the mixed-grass prairies. Short-grass tropical savannas characterized by the presence of small trees on the uplands are sometimes called tropical steppes.

Steppe grass supports much of the natural grazing of the world both fixed and nomadic, and do-



Fig. 1. Herd of bison in Niobrara Wildlife Refuge, Nebraska. Short steppe grass of the foreground is typical of the Great Plains of Interior United States. (U.S. Fish and Wildlife Service, Department of the Interior)



Fig. 2. Short-grass tropical savanna with scattered trees in Africa. The native animals are the gazelle, eland, and hartebeest. (American Museum of Natural History)

mestic animals have largely replaced the native grazing fauna, notably the North American bison. It is believed that most domestic grazing animals, especially cattle and horses, came originally from the Eurasian steppe grasslands. *See* PRAIRIE; VEGETATION ZONES (WORLD). [C.M.D.]

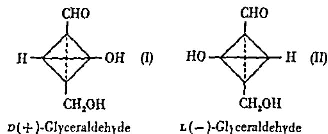
Stereochemistry

The study of the spatial arrangement of atoms in molecules and the chemical and physical consequences of such arrangement. Two important types of organic molecules are included in the general class of stereoisomers: geometrical and optical isomers. *See* ISOMERISM, MOLECULAR; OPTICAL ACTIVITY. Such isomers differ from each other, not by having different structural groupings of the same elements, but by having different arrangements of their constituent atoms in space, while having the same structural features; that is, in each isomer every atom is bonded to the same immediate neighbors.

Once the [illegible] molecular d[illegible] van't Hoff a[illegible] cenus, 1877) were understood, a number of experimental problems immediately posed themselves and have continued to occupy prominent positions in the field of organic chemistry: the interrelation of optically active centers in the same and in different molecules (relative and absolute configurations), the interconversion of enantiomorphs (*see* RACEMIZATION), the conversion of one configuration into its opposite during replacement of a substituent (Walden inversion), the separation of enantiomorphs from each other, the production of optically active substances from optically inactive sources (*see* ASYMMETRIC SYNTHESIS), and the synthesis of a specific stereoisomer from comparatively simple starting materials without recourse to multiple isomer separations and more than one resolution.

Absolute and relative configurations. The importance of optically active substances of physiological significance (carbohydrates and amino acids) demands specific knowledge of configurational (spatial) relationships among them. Histor-

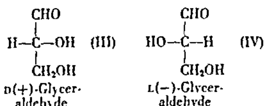
ically the dextro form, or (+), of glyceraldehyde was selected by Emil Fischer as a configurational standard and arbitrarily was assigned the right-handed, or D, configuration (I). The levo form, or (−), thus is left-handed and designated L (II).



D(+)-Glyceraldehyde L(−)-Glyceraldehyde

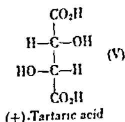
The central carbon is asymmetric and is represented by the formalized tetrahedron, the corners of which represent the bonding positions of substituents

It is convenient to replace such representations by projection formulas by compressing the tetrahedra into the plane of the paper. Such flattening at once imposes certain restrictions on the handling of a projection formula to avoid confusion and retain explicitness: it must never be "removed" from the plane of the paper, it must never be "rotated" in this plane through any but integral multiples of 180°; and the "exchange" of any opposite pair of substituents is equivalent to transforming it into its mirror image. The projection formulas (III) and (IV) represent the enantiomorphic glyceraldehydes.



D(+)-Glyceraldehyde L(−)-Glyceraldehyde

Because of confusion arising from different systems of configurational assignments to polycentered asymmetric compounds (for example, α-amino acids are classified by the configuration of the lowest-numbered asymmetric carbon, whereas carbohydrates are classified according to the configuration of the highest-numbered asymmetric carbon, and hydroxy acids such as tartaric are equivocal), current usage requires that each asymmetric carbon be designated D or L in a projection formula with carbon-1 uppermost, using the International Union of Chemistry system of numbering. Thus (+)-tartaric acid (V) is 2 D,3-L-dihydroxybutanedioc acid.



(+)-Tartaric acid

Originally the configurations represented by (III), (IV), and (V) were mutually consistent, although arbitrary. Recently J. M. Bijvoet has determined the actual or absolute configurations by crystallographic study of (+)-tartrate salts, finding them to agree with the arbitrary assignment. Thus the absolute configuration of any substance unequivocally relatable to the tartaric acids is practically determinable.

A long-standing suggestion that anomalous rotatory dispersion might be useful in determining configurations has recently been realized by Carl Djerassi, who has employed it effectively for configurational determinations in such complex systems as steroids and terpenes. In consequence, other methods for configurational determination through optical properties become chiefly of historical or theoretical interest only.

The term relative configuration has been used somewhat loosely to indicate configuration relative to glyceraldehyde, relative to some other substance, or relative to another asymmetric center in the same molecule. Since absolute configurational determination is now possible, relative configuration is best confined to intramolecular configurational relationships and those in other diastereoisomeric forms of the same substance. Relative configurations may be determined without reference to glyceraldehyde or even to optical properties. For example the meso and (±) configurations of stilbene dibromide are determinable from a knowledge of reaction mechanisms. *See* CONFORMATIONAL ANALYSIS.

On the other hand, in order to know whether a configurational inversion has occurred in the course of a reaction, explicit information concerning the configurations of starting material and product is required For example, a difference in reaction conditions for the conversion of optically active 1-phenylethanol to 1-phenylethyl chloride, results in the formation of either enantiomorph from the same source When the configurations are known, information as to the stereochemistry of the replacement of hydroxyl by chlorine becomes available and may be applied to other similar substitutions, regardless of optical consequences and without reference to absolute configurations.

Resolution. The isolation of either or both enantiomorphs from a racemic mixture is termed resolution It can be achieved only with the assistance of an asymmetric reagent, since each enantiomorph will behave identically in a symmetrical environment, chemical or physical.

The more general procedure involves the readily reversible conversion of both enantiomorphs into a pair of diastereoisomers with as widely different physical properties as practicable by reaction with a single pure optical isomer Thus, a (±) amine reacts with a (+) acid to give two diastereomeric salts:



and

$$(-)\text{amine}\cdot(+)\text{acid}$$

Since these are not enantiomorphic, they will have different properties and may be separated by fractional recrystallization. The salts can then be converted to resolved amines by destroying the salt with a strong base. Many modifications of this technique are possible

Less frequently, resolution is effected kinetically, by taking advantage of differences in reaction rates of enantiomorphs with an asymmetric reagent. Enzymatic resolutions, in which an enzyme *preferentially* destroys one enantiomorph, and *chromatography* on asymmetric adsorbents, which binds one enantiomorph *more strongly*, constitute examples of this type of resolution.

The determination of the total number of theoretically possible optical isomers is principally of academic interest. In the general case of a substance with *n* different asymmetric centers, this number is 2^n; the presence of identical centers, or of structural features which disallow certain configurations, will decrease this number.

Stereospecific synthesis. The multiplicity of possible stereoisomers for a natural product such as a steroid, terpene, or alkaloid has focused considerable attention on reactions which produce but one of two or more stereoisomers To this end, the stereospecificity of many simple and complex reactions is of paramount interest The study of steric effects caused by the bulk, polarity and consequent interaction with neighboring groups, conformational requirements, and intimate details of reaction mechanisms has enabled the total synthesis of stereochemically complex substances without the necessity of isomer separations and with but one resolution at or near the end of a long reaction series

Notable among the simpler examples of stereospecific reactions are the following: kinetically apparent second-order displacements involving Walden inversion, displacements in which configuration is retained because of mechanistic geometry or participation of groups favorably located with respect to the reaction center, cis and trans additions to π bonded systems and cis and trans eliminations which produce π bonds, the Diels-Alder diene reaction In addition, certain molecular rearrangements are stereospecific: thus the Wagner-Meerwein rearrangement often effects complete inversion at the reaction center, and the Curtius, Hofmann, Schmidt and Wolff rearrangements occur with virtually complete retention of the configuration of the migrating group *See* COORDINATION CHEMISTRY [W.R.V.]

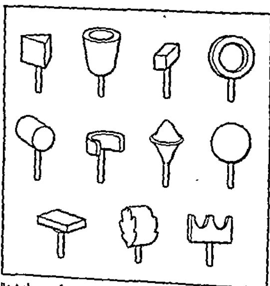
Bibliography: H Gilman et al. (eds.), *Organic Chemistry*, 2d ed, vol 1, 1943; M. S Newman, *Steric Effects in Organic Chemistry*, 1956.

Stereognosis

The recognition or identification of objects exclusively through handling them Basically, this ability depends on the accuracy with which the cutaneous and kinesthetic senses can combine their discriminative capacities. *See* KINESTHETIC SENSATION; TOUCH.

A necessary condition for the skin to appreciate form of objects or texture of surfaces is that there be relative motion involved; an object laid statically on the skin cannot be judged as to shape If, however, the finger tips move successively over two glass surfaces, one smooth, the other etched to provide slight eminences no greater than a thousandth of a millimeter high, the textural difference can be detected. Cloth feelers, who work in weaving mills, become so familiar with such cues as to make highly accurate discriminations on the basis of a single, brief manipulation.

To identify complexly formed solids it is obviously necessary to integrate a series of impressions, the sensitive finger tips being passed over all salient features of the object, pressing and squeezing as well as hefting, which provides the kinesthetic component of the judgment.



Best shapes for control knobs. These shapes have been found to be the most easily recognized by feel alone.

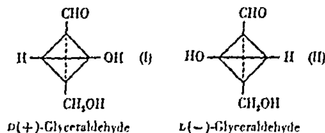
The stereognostic capacity has been put to work in the shape coding of airplane controls. Levers may thus be recognized without the aid of sight. Experiments have also been performed on letters and figures, the finger tip being the skin area under test. If six symbols were to be selected, for example, for tactual coding of a keyboard, they would be C, I, O, 7, L, V. These are speedily and correctly identified by touch alone [F.A C.]

Bibliography: M. A Wenger, F. N. Jones, and M H. Jones, *Physiological Psychology*, 1956.

Stereophonic sound

...ources are substantially retained; that is, the listener hears the sound in auditory perspective.

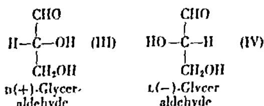
ically the dextro form, or (+), of glyceraldehyde was selected by Emil Fischer as a configurational standard and arbitrarily was assigned the right-handed, or D, configuration (I). The levo form, or (−), thus is left-handed and designated L (II).



D(+)-Glyceraldehyde L(−)-Glyceraldehyde

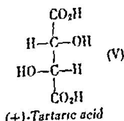
The central carbon is asymmetric and is represented by the formalized tetrahedron, the corners of which represent the bonding positions of substituents.

It is convenient to replace such representations by projection formulas by compressing the tetrahedra into the plane of the paper. Such flattening at once imposes certain restrictions on the handling of a projection formula to avoid confusion and retain explicitness: it must never be "removed" from the plane of the paper, it must never be "rotated" in this plane through any but integral multiples of 180°, and the "exchange" of any opposite pair of substituents is equivalent to transforming it into its mirror image. The projection formulas (III) and (IV) represent the enantiomorphic glyceraldehydes.



D(+)-Glyceraldehyde L(−)-Glyceraldehyde

Because of confusion arising from different systems of configurational assignments to polycentered asymmetric compounds (for example, α-amino acids are classified by the configuration of the lowest-numbered asymmetric carbon, whereas carbohydrates are classified according to the configuration of the highest-numbered asymmetric carbon, and hydroxy acids such as tartaric are equivocal), current usage requires that each asymmetric carbon be designated D or L in a projection formula with carbon-1 uppermost, using the International Union of Chemistry system of numbering. Thus (+)-tartaric acid (V) is 2-D,3-L-dihydroxybutanedioc acid.



(+)-Tartaric acid

Originally the configurations represented by (III), (IV), and (V) were mutually consistent, although arbitrary. Recently J. M. Bijvoet has determined the actual or absolute configurations by crystallographic study of (+)-tartrate salts, finding them to agree with the arbitrary assignment. Thus the absolute configuration of any substance unequivocally relatable to the tartaric acids is practically determinable.

A long-standing suggestion that anomalous rotatory dispersion might be useful in determining configurations has recently been realized by Carl Djerassi, who has employed it effectively for configurational determinations in such complex systems as steroids and terpenes. In consequence, other methods for configurational determination through optical properties become chiefly of historical or theoretical interest only.

The term relative configuration has been used somewhat loosely to indicate configuration relative to glyceraldehyde, relative to some other substance, or relative to another asymmetric center in the same molecule. Since absolute configurational determination is now possible, relative configuration is best confined to intramolecular configurational relationships and those in other diastereoisomeric forms of the same substance. Relative configurations may be determined without reference to glyceraldehyde or even to optical properties. For example the meso and (±) configurations of stilbene dibromide are determinable from a knowledge of reaction mechanisms See CONFORMATIONAL ANALYSIS

On the other hand, in order to know whether a configurational inversion has occurred in the course of a reaction, explicit information concerning the configurations of starting material and product is required For example, a difference in reaction conditions for the conversion of optically active 1-phenylethanol to 1-phenylethyl chloride, results in the formation of either enantiomorph from the same source When the configurations are known, information as to the stereochemistry of the replacement of hydroxyl by chlorine becomes available and may be applied to other similar substitutions, regardless of optical consequences and without reference to absolute configurations

Resolution. The isolation of either or both enantiomorphs from a racemic mixture is termed resolution It can be achieved only with the assistance of an asymmetric reagent, since each enantiomorph will behave identically in a symmetrical environment, chemical or physical

The more general procedure involves the readily reversible conversion of both enantiomorphs into a pair of diastereoisomers with as widely different physical properties as practicable by reaction with a single pure optical isomer. Thus, a (±) amine reacts with a (+) acid to give two diastereomeric salts:


(+)amine·(+)acid

and

(−)amine•(+)acid

Since these are not enantiomorphic, they will have different properties and may be separated by ... llization The salts can then be ... ying the salt ... of this technique are possible.

Less frequently, resolution is effected kinetically, by taking advantage of differences in reaction rates of enantiomorphs with an asymmetric reagent. Enzymatic resolutions, in which an enzyme preferentially destroys one enantiomorph, and chromatography on asymmetric adsorbents, which binds one enantiomorph more strongly, constitute examples of this type of resolution.

The determination of the total number of theoretically possible optical isomers is principally of academic interest. In the general case of a substance with n different asymmetric centers, this number is 2^n; the presence of identical centers, or of structural features which disallow certain configurations, will decrease this number

Stereospecific synthesis. The multiplicity of possible stereoisomers for a natural product such as a steroid, terpene, or alkaloid has focused considerable attention on reactions which produce but one of two or more stereoisomers To this end, the stereospecificity of many simple and complex reactions is of paramount interest. The study of steric effects caused by the bulk, polarity and consequent interaction with neighboring groups, conformational requirements, and intimate details of reaction mechanisms has enabled the total synthesis of stereochemically complex substances without the necessity of isomer separations and with but one resolution at or near the end of a long reaction series

Notable among the simpler examples of stereospecific reactions are the following: kinetically apparent second-order displacements involving Walden inversion, displacements in which configuration is retained because of mechanistic geometry or participation of groups favorably located with respect to the reaction center, cis and trans additions to π-bonded systems and cis and trans eliminations which produce π bonds, the Diels-Alder diene reaction In addition, certain molecular rearrangements are stereospecific, thus the Wagner-Meerwein rearrangement often effects complete inversion at the reaction center, and the Curtius, Hofmann, Schmidt, and Wolff rearrangements occur with virtually complete retention of the configuration of the migrating group See COORDINATION CHEMISTRY [W R V.]

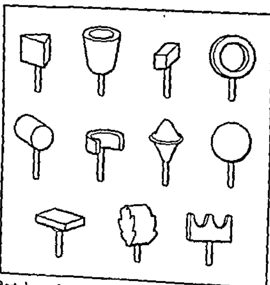
Bibliography H Gilman et al (eds), *Organic Chemistry*, 2d ed, vol 1, 1943; M. S. Newman, *Steric Effects in Organic Chemistry*, 1956

Stereognosis

The recognition or identification of objects exclusively through handling them Basically, this ability depends on the accuracy with which the cutaneous and kinesthetic senses can combine their discriminative capacities. See KINESTHETIC SENSATION; TOUCH.

A necessary condition for the skin to appreciate form of objects or texture of surfaces is that there be relative motion involved; an object laid statically on the skin cannot be judged as to shape. If, however, the finger tips move successively over two glass surfaces, one smooth, the other etched to provide slight eminences no greater than a thousandth of a millimeter high, the textural difference can be detected. Cloth feelers, who work in weaving mills, become so familiar with such cues as to make highly accurate discriminations on the basis of a single, brief manipulation.

To identify complexly formed solids it is obviously necessary to integrate a series of impressions, the sensitive finger tips being passed over all salient features of the object, pressing and squeezing as well as hefting, which provides the kinesthetic component of the judgment.



Best shapes for control knobs These shapes have been found to be the most easily recognized by feel alone

The stereognostic capacity has been put to work in the shape coding of airplane controls. Levers may thus be recognized without the aid of sight. Experiments have also been performed on letters and figures, the finger tip being the skin area under test. If six symbols were to be selected, for example, for tactual coding of a keyboard, they would be C, I, O, 7, L, V. These are speedily and correctly identified by touch alone [F.A.G.]

Bibliography: M. A. Wenger, F. N. Jones, and M H Jones, *Physiological Psychology*, 1956.

Stereophonic sound

Sound which is reproduced or transmitted in such a manner that the spatial relations of the original sound sources are substantially retained; that is, the listener hears the sound in auditory perspective.

Stereophonic sound-reproduction systems, introduced on a wide scale in the 1950s for motion pictures, disk phonographs, and magnetic tape recorders, are based on the fact that a person with normal hearing can determine the direction from which a sound is coming by distinguishing differences in arrival times of sound waves at his two ears. This ability is known as binaural hearing; for details, see HEARING.

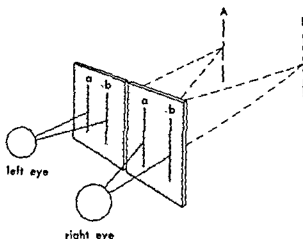
Two or more separate sound-recording and reproducing channels are necessary to provide reproduction of sound sources in auditory perspective. For a detailed discussion of the recording and reproducing techniques used to provide stereophonic sound, see DISK RECORDING; MAGNETIC RECORDING; OPTICAL RECORDING; SOUND REPRODUCTION SYSTEMS, ELECTRICAL. *See also* BINAURAL SOUND; HIGH FIDELITY. [K.W.P.]

Stereoscopy

The phenomenon of simultaneous vision with the two eyes wherein there arises a visual experience of the third dimension, that is, a vivid perception of the relative distances of objects in space. In this experience the observer seems to see the space between the objects located at different distances from the eyes. The stereoscopic effect is so unique that it cannot be easily described to one who does not possess it. Stereopsis, or stereoscopic vision, provides the individual with the most acute sense of relative depth and is of vital importance in visual tasks requiring the precise location of objects.

Stereopsis is believed to have an innate origin in the anatomic and physiologic structures of the retinas of the eyes and the visual cortex. It is present in normal binocular vision because the two eyes view objects in space from two points, so that the retinal image patterns of the same object points in space are slightly different in the two eyes. The stereoscope, with which different pictures can be presented to each eye, demonstrates the fundamental difference between stereoscopic perception of depth and the conception of depth and distance from the monocular view. In the illustration, each of the two eyes views a pair of vertical lines A and B drawn on cards. The separation of these lines for the right eye is greater than that for the left eye. If the difference in separation is not too great, the images of the lines fuse when the two targets are observed by the two eyes. There is almost an immediate stereoscopic spatial experience that the two lines are located in space, line B being definitely more distant than line A. No hint of this spatial experience occurs in the observation of each target alone. It is this difference between the images in the two eyes that provides the stimulus for the emergence of the stereoscopic experience.

Stereoscopic acuity and the presence of stereopsis may be tested by several methods or instruments such as the ordinary hand stereoscope and similar devices, vectograph targets, the Howard-Dolman peg test, the Verhoeff stereopter, the Hering falling bead test, pins with colored heads stuck in a board to test for near vision, and afocal meridional magnifying lenses placed before one eye while the subject views a special field (leaf room or tilting table). *See* VECTOGRAPH; VISION. [K.N.O.]



Stereoscopic vision.

Bibliography: A. Ames, Jr., Binocular vision as affected by relations between uniocular stimulus-patterns in commonplace environments, *Am J Psychol.*, 59:333, 1946; K. N. Ogle, *Researches in Binocular Vision*, 1950; K. N. Ogle, Present status of our knowledge of stereoscopic vision, *AMA Arch. Ophthalmol.*, 60:755, 1958.

Stereospondyli

A group of Triassic amphibians, the last survivors of the superorder Labyrinthodontia, defined by the formation of the vertebral centra from the intercentrum alone. They were degenerate in many regards; the skull and body were flattened, skeletal ossification was much reduced, and the limbs were small. The stereospondyls were purely aquatic and apparently incapable of progression on land. The development of a double occipital condyle and of large paired palatal vacuities are among their notable cranial characters. There were several families, the Capitosauridae preserving a normal skull pattern, the Metopsauridae having the eyes far forward in the skull, and the Brachyopidae exhibiting a very broad and short skull. *See* LABYRINTHODONTIA. [A.S.R.]



Dorsal view of Triassic stereospondyl *Buettneria*, showing orbits near front of skull, 1/30 natural size. (*After Sawin*)

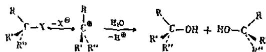
Steric effect (chemical reaction)

The influence of the spatial configuration of reacting substances upon the rate, nature, and extent of reaction. The sizes and shapes of atoms and molecules, the electrical charge distribution, and the geometry of bond angles influence the courses of chemical reactions.

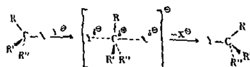
The steric course of organochemical reactions is greatly dependent on the mode of bond cleavage and formation, the environment of the reaction site, and the nature of the reaction conditions (reagents, reaction time, and temperature). The effect of steric factors is best understood in ionic reactions in solution. The nucleophilic substitution reaction at a saturated carbon atom can serve as an illustration.

Saturated nucleophilic substitution. While the reacting carbon atom is in an electron-deficient state in the transition state (state of highest energy, somewhere between starting material and product) in a nucleoph[illegible]

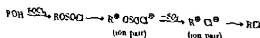
The former mode of reaction, a solvolysis or S_N1 process, converts a tetrahedral carbon into a solvated planar carbonium ion intermediate, and hence, leads to racemization (randomization of configuration); for example,



The second reaction path proceeds by a simultaneous rupture of the old bond and creation of a new one, either by inversion of configuration, a Walden inversion or S_N2 process, for example,

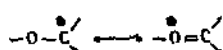


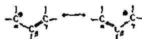
or in a few cases by a front-side displacement of part of the substituent already present, an S_Ni process, and hence, retention of configuration, for example,

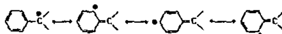


These substitution reactions are highly solvent-dependent, for example, *tert*-butyl chloride undergoes solvolysis close to 500,000 times faster in water, a solvent of high ionizing power, than in ethanol, a solvent of low dielectric properties. The nature of R, R', and R'' is one of the factors determining which of the above pathways a compound prefers for its substitution. The larger the size of these three groups, the greater is the tendency to relieve steric strain by extrusion of X, and thus, the more the need for an S_N1 process. The smaller the size of the environment, the greater is the accessibility of the reagent from the back side, and thus, the more tendency toward an S_N2 process. As a consequence, tertiary compounds undergo racemization readily, whereas primary systems prefer inversion.

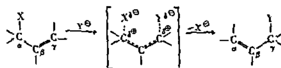
Both the rate and the steric course of an ionic displacement may depend often on the ability of groups adjacent to the reaction site to accommodate a positive charge. Substitution of α-halo ethers and allyl or benzyl halides occurs much faster than a similar reaction of unsubstituted halides because of the intermediacy of the following stabilized cations or cationlike transition states:



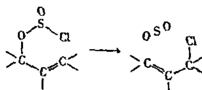




In view of the charge distribution over more than one atom in these cases, sometimes the incoming reagent forms a bond at a site different from that of the leaving group. As a consequence, a S_N2' process

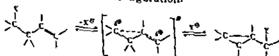


or a S_Ni' path

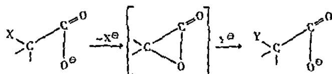


may result. Both processes yield retention of configuration; that is, the orientation of the new substituent on its carbon atom is identical with that originally held by the former functional group on a site two carbon atoms removed.

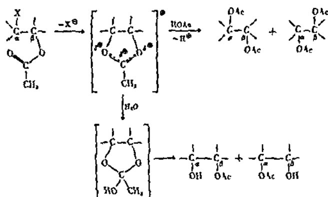
In the presence of participating neighboring groups, even solvolyses can lead to retention of configuration. Certain rigidly held homoallyl systems undergo substitution at a site three carbons removed from the position of the leaving group, but with retention of configuration.



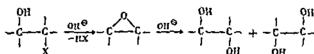
Double inversion is responsible for the retained configuration of the products of solvolysis of α-halo acid salts



Solvolysis of *trans*-β-acetoxy systems in nonaqueous media leads to trans products. In the presence of water, cis compounds are obtained:

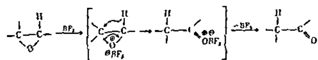


Base treatment of *trans*-halohydrins leads to *trans*-vicinal glycols. The intermediate epoxide is isolable. Although ring opening of the latter may yield two different trans products, the diaxial one is formed preferentially in cycloalkane cases

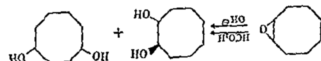


Rearrangement. Migration of neighboring groups toward the reaction site, resulting in skeletal rearrangements, is a common occurrence. Both the internal displacement of the leaving group by the migrating group and the subsequent external displacement of the migrating group by the solvent or added reagent proceed in a trans sense, that is, by back-side approach. Thus, the over-all steric consequence of one migration sequence is retention of configuration.

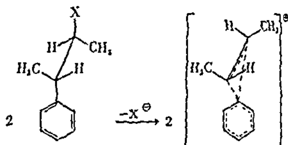
There are several examples of 1,2-hydride migration, for example,

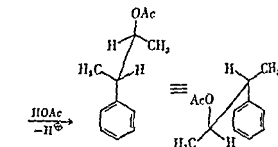


Transannular hydride shifts are quite similar in nature

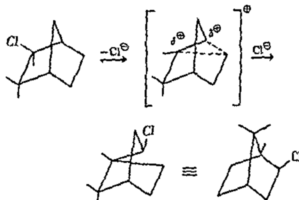


The 1,2 migration of phenyl groups proceeds by way of fairly stable phenonium ions

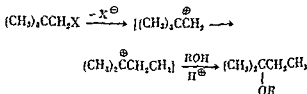




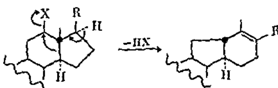
The Wagner-Meerwein rearrangement of saturated neighboring groups shows directional effects similar to those of the above migrations; for example, the conversion of camphene hydrochloride to bornyl chloride is stereospecific, with retention of configuration.



Neopentyl halides solvolyze to tertiary amyl derivatives



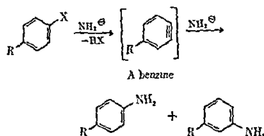
Cyclohexane systems with equatorial leaving groups may undergo contraction to five-membered rings.



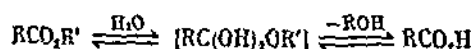
Organic compounds possessing potential leaving groups at the bridgehead of sm clic ring systems undergo substitution nly s

gishly. In the absence of ready access at the back side of the reaction center, the S_N2 pathway is excluded The inability of the compounds to form planar carbonium ions precludes a S_N1 route. However, displacements do occur slowly at elevated temperatures, presumably via nonplanar cations

Unsaturated nucleophilic substitution. Nucleophilic substitution reactions at unsaturated carbon atoms can take place by two possible mechanistic routes, an elimination-addition process and an addition-elimination scheme. The former route is best illustrated by the transformation of aromatic halides into anilines

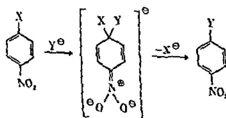


The latter is encountered in the interconversion of carboxylic acids and their derivatives, for example,

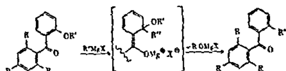


Because the central carbon atom has greater steric requirements in the reaction intermediate than in the starting material, the reaction velocity is strongly dependent on the size and number of neighboring groups, that is, an increase in the bulkiness of R is reflected in a decrease of the rate of the reaction

The addition-elimination mechanism is portrayed also by the aromatic nucleophilic substitution reaction, for example,

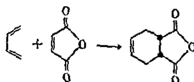


In order to be able to stabilize the reaction intermediate, the all important nitro group must be coplanar with the benzene ring. As a consequence, ortho substituents, which may block this steric requirement, retard the reaction rate. Unusual aromatic nucleophilic substitutions have been observed in cases where steric hindrance by ortho substituents has prevented addition to aromatic ketones to occur, for example,



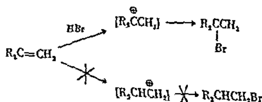
Addition reactions. The steric course of addition reactions at unsaturated sites depends largely on the reagent. Catalytic hydrogenation, a nonhomogeneous process of undetermined mechanism, occurs in a cis manner. In the absence of any steric interference, it leads to thermodynamically stable products In the presence of steric hindrance, the two new hydrogen atoms are usually introduced on the least hindered side of the unsaturated compounds. However, sometimes some bulky polar groups actually aid, rather than retard, adsorption of the catalyst on their side of the reducing compound, thereby leading to products of opposite configuration

The oxidation of olefins to vicinal glycols by permanganate salts or osmium tetroxide also proceeds in a cis fashion and also involves the least-hindered side of the reacting substrate. The Diels-Alder reaction behaves similarly, for example,

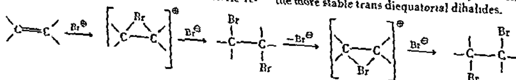


All addition processes, during which two new bonds are formed more or less simultaneously, yield cis adducts

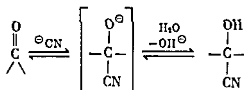
Ionic add
occur in a
such as to
ions (Mark addition), for example,



Halogen addition to cyclic olefins leads to trans diaxial dihalides, which, on standing, isomerize to the more stable trans diequatorial dihalides.

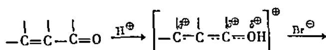


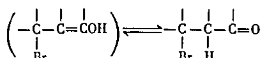
Ionic addition reactions of carbonyl compounds follow a steric course very similar to those of olefins. However, the reagents are mostly nucleophilic, and some reactions are equilibrium processes, for example,

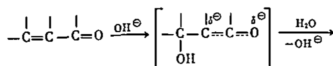


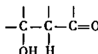
The orientation of attack and the reaction rate are governed by the environment of the carbonyl group.

Addition reactions of conjugated carbonyl systems can occur through cation as well as anion intermediates, buy they uniformly place the nucleophilic part of the reagents on the β-carbon atoms, for example,



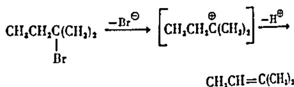




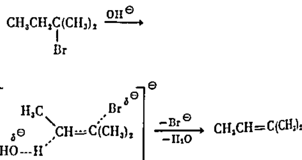


The reaction of carbonyl compounds as enol or enolate anions with electrophilic reagents can take two different courses If the process is kinetically controlled, the electrophile, a proton, halonium ion, or others, interacts with the substrate on its least-hindered side. However, if the reaction is thermodynamically controlled, then, independent of mechanism, the most stable product is obtained.

Elimination reactions. Elimination reactions can be carried out by pyrolysis of esters, halides, or amine oxides in the liquid or vapor phase These eliminations always involve a rupture of vicinal cis bonds. Alternatively, similar cleavage processes can be made to occur ionically in solution, in which case they proceed in a trans fashion. The direction of elimination depends greatly on the molecularity of the process, as well as on the sizes of the leaving group and the attacking base. The two-step, unimolecular cleavage, an $E1$ process, leads predominantly to the more substituted, hence more stable, olefins, for example,

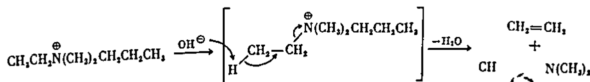


The one-step, bimolecular elimination (an $E2$ process) of neutral compounds yields similar products, for example,

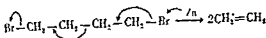


However, an $E2$ reaction on positive ions, ammonium or sulfonium salts, affords the less substituted olefin in preponderant yield (for an example, see below)

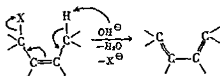
Reactions leading to the more stable products are said to follow the Saytzeff rule, whereas those yielding less stable olefins obey the Hofmann rule Because the transition state in the $E2$ reaction is of lowest energy when all atoms involved in the elimination are in a plane, the fastest rates among cyclic compounds are encountered in the cases which permit a diaxial alignment of vicinal substituents



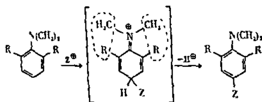
Many ionic elimination reactions are known, which involve the rupture of more than two bonds, for example,



The $E2'$ processes, eliminations of two groups on carbon atoms separated by an olefinic linkage, appear to be cis in nature



Electrophilic substitution. Steric factors have a fair control over the course of the aromatic electrophilic substitution reaction In reactions of compounds containing ortho-para directing substituents, the p/o product ratio is usually greater than 1, and increases with the size of the substituent and that of the reacting species The rate-accelerating participation of electron-donating groups, located ortho or para to the incoming substituent, in stabilizing the transition state is greatly diminished in the presence of bulky ortho neighbors which would prevent the groups from attaining coplanarity with the benzene, for example.



See CHELATION, CONFORMATIONAL ANALYSIS; ORGANIC CHEMICAL SYNTHESIS; ORGANIC REACTION MECHANISM, STEREOCHEMISTRY. [E W]

Bibliography W Klyne (ed), *Progress in Stereochemistry*, vol. 1, 1954, M S Newman (ed), *Steric Effects in Organic Chemistry*, 1956.

Sterility

In man, relative or absolute inability of either the male or the female to reproduce In the male, sterility may or may not be associated with impotence the inability to perform the sexual act. Absent or abnormal testes, the production of low numbers of or imperfectly formed sperm, and obstruction of the passage of sperm are the most common causes in the male Congenital defects, hormonal imbalances, or chronic disease may contribute to sterility in either sex.

In women, the most common causes are related to obstruction of the Fallopian tubes and abnormalities of the cyclic changes which occur in the uterus and other genital organs

In any case of continuing failure to achieve pregnancy both husband and wife should be thoroughly examined by a specialist in the field. In most cases relatively minor adjustments will suffice; in a few cases the causes cannot be remedied by present means *See* REPRODUCTIVE SYSTEM. [E G ST.]

Sterilization

An act of destroying all forms of life on and in an object. A *substance* is sterile, from a microbiological point of view, when it is free of all living microorganisms. Sterilization is used principally to prevent contamination which often causes the spoilage of food and other substances and to prevent the transmission of diseases, by destroying microbes [illegible] diseases in man and animals

are killed, they generally die at a constant rate under specified environmental conditions. If the logarithm of the number of survivors is plotted against time, the resulting curve will produce a straight line

When testing a substance for sterility, care must be taken to employ appropriate techniques A bacterial cell is considered to be killed when it is no longer capable of reproducing itself under suitable environmental conditions. If an inadequate medium is employed to subculture the treated bacteria, the substance being tested may be wrongly considered to be sterile.

By far the most resistant of all forms of life, to both physical and chemical killing agents, are some of the bacterial endospores (*see* BACTERIAL ENDOSPORES) If they did not exist, sterilization of such materials as bacteriological media and equipment, hospital supplies, and canned foods would be much simpler.

Heat sterilization. This is the most common method of sterilizing bacteriological media, foods, hospital supplies, and other substances. Either moist heat (hot water or steam) or dry heat can be employed, depending upon the nature of the substance to be sterilized Moist heat is also used in pasteurization, which is not considered a true sterilization technique because all microorganisms are not killed, certain pathogenic organisms and other undesirable bacteria are destroyed (*see* PASTEURIZATION).

Moist-heat sterilization. Some bacterial endospores are capable of surviving several hours at 100°C. Therefore, for moist-heat sterilization, an autoclave, pressure cooker, or retort, with steam under pressure, is required to achieve higher temperatures

Most bacteriological media are sterilized by autoclaving with steam at 121°C, with 15 lbs pressure, for 20 min or more, depending upon the volume of material being heated. Some spores are capable of surviving moist heat equivalent to at least 7 min at 121°C.

Steam under atmospheric pressure in an Arnold sterilizer is sometimes employed for specialized bacteriological media that are easily heat dam-

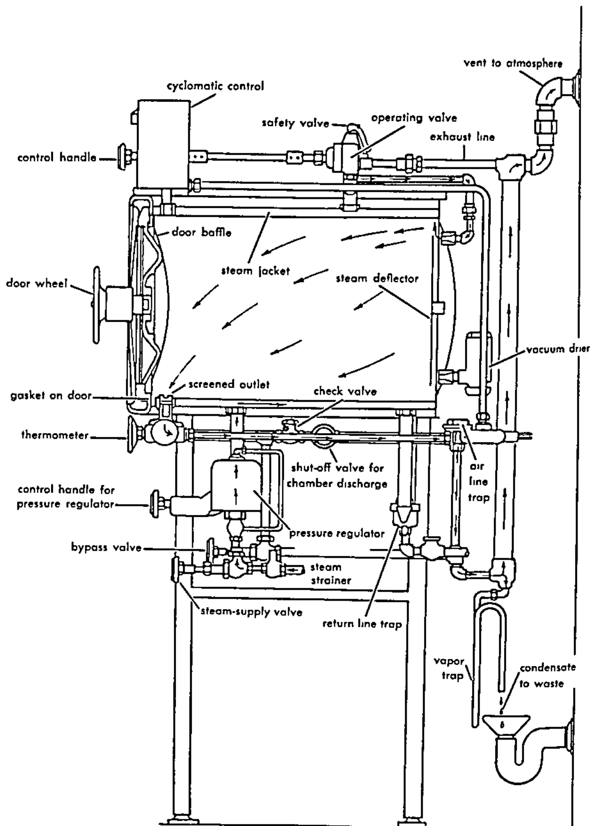


Fig. 1. Diagrammatic sketch of an autoclave. (*American Sterilizer Company*)

aged. Because many bacterial spores survive this treatment, it is obviously inadequate to ensure sterility.

Tyndallization. The food or medium is steamed for a few minutes at atmospheric pressure on three or four successive occasions, separated by 12- to 18-hour intervals of incubation at a favorable growing temperature. In theory the intervals of incubation allow any surviving bacterial spores to germinate into more heat-sensitive, vegetative cells which then would be killed during the next heat treatment. However, spores like vegetative cells may require special conditions such as an appropriate medium or proper oxygen tension or proper

temperature to germinate and reproduce. These conditions may not be realized during the intervals between heat treatment and no matter how often the steaming is repeated ungerminated spores may survive and eventually germinate when conditions have been changed. The survival of ungerminated spores reduces the effectiveness of this method and it has been supplanted by other methods.

Hot air sterilization Glassware and other heat resistant materials which need to be dry after treatment are usually sterilized in a hot-air sterilizer. Dry sterilization requires heating at higher temperatures and for longer periods of time than does sterilization by steam under pressure. A temperature of 160–165°C for at least 2 hours is generally employed in hot-air sterilization. Dry heat kills the germs through denaturation of protein which may involve oxidative processes.

Radiation sterilization. Many kinds of radiations are lethal, not only to microorganisms but to other forms of life These radiations include both high energy particles as well as portions of the electromagnetic spectrum The mechanism of the lethal action of these radiations is not entirely clear It may involve a direct energy absorption at some vital part of the cell (direct target theory), or the production of highly reactive, ionized, free radicals near some vital part of the cell (indirect effect) Bacterial endospores are relatively resistant to all types of radiation See RADIATION BIOCHEMISTRY

Ultraviolet radiation. Radiant energy in the ultraviolet region of the spectrum is highly bactericidal, especially at wavelengths of approximately 2650 angstroms Lamps which generate ultraviolet radiation in this region are useful for the sterilization of air and smooth surfaces Ultraviolet rays have very low penetrative capacity, since even a thin layer of glass absorbs a high percentage of the rays Some irradiated cells that [illegible] dead may [illegible]

Gamma [illegible] netic radia[illegible] penetrative [illegible] and their energy is dissipated in the production of ionized particles from the material being irradiated Radioactive isotopes, such as cobalt 60, are a common source of gamma rays. Gamma irradiation of foods has received much attention as a means of sterilizing foods without cooking them Sterilization requires a radiation dose of approximately 5,000,000 rads (5×10^{8} ergs of energy absorbed per gram)

Cathode rays These high-speed electrons (beta rays) may be generated with various types of electron accelerators This type of radiation has relatively low penetrative capacity, depending upon the energy level of the emitted electron beam. Cathode rays sterilize in a manner identical with gamma rays and without significantly raising the temperature of the material being irradiated. They have received some application in the sterilization of surgical supplies, drugs, and in the experimental sterilization of foods.



Fig 2 Cobalt-60 furnace designed for experimental food sterilization.

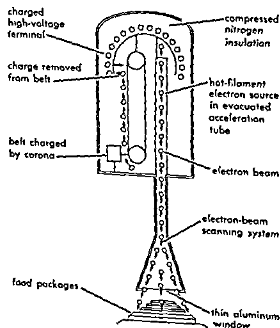


Fig 3. Von de Graaff Generator principle, packaged products sterilized by means of high-intensity electron beam which penetrates packages and sterilizes contents. (From W. M. Urbain, Food Eng, vol. 25, February, 1953)

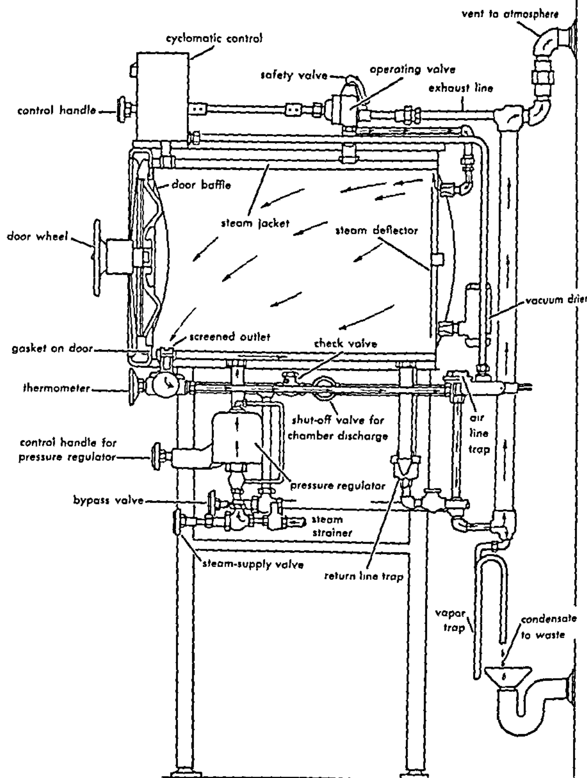


Fig. 1. *Diagrammatic sketch of an autoclave (American Sterilizer Company)*

aged. Because many bacterial spores survive this treatment, it is obviously inadequate to ensure sterility.

Tyndallization. The food or medium is steamed for a few minutes at atmospheric pressure on three or four successive occasions, separated by 12- to 18-hour intervals of incubation at a favorable growing temperature. In theory the intervals of incubation allow any surviving bacterial spores to germinate into more heat-sensitive, vegetative cells, which then would be killed during the next heat treatment. However, spores like vegetative cells may require special conditions such as an appropriate medium or proper oxygen tension or proper

benzene, dichloroethane, and *n*-butyl chloride, as employed by S. H. Hutner. [C.F.N.]

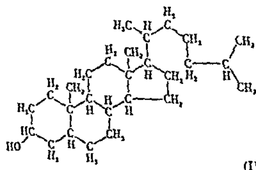
Bibliography: M. J. Pelczar and R. D. Reid, *Microbiology*, 1958, G. F. Reddish (ed.), *Antiseptics, Disinfectants, Fungicides, and Chemical and Physical Sterilization*, 2d ed., 1957.

Steroid

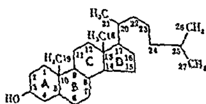
One of a group of widely occurring natural products. The steroids are critically important to plant and animal life. Many steroids, like sterols, bile acids, sex hormones (androgens and estrogens), adrenal cortex steroids (adrenal cortex hormones), and cardiac active principles (digitoxigenin), exert profound physiological effects. Steroids are used on an ever increasing scale in medicine for the treatment of disease. New types of steroids and new methods of production are being developed. In general, steroids used in the field of medicine are prepared by a combination of chemical and microbiological methods.

Structure. The steroids are solids, colorless, and for the most part, saturated compounds (contain few double bonds —C=C—). They contain a complicated ring structure, the cyclopentanoperhydrophenanthrene ring system or a close modification of this. The substances within this family differ in the chemical substituents on the ring system, which is also called the nucleus, and on the side chain attached to the nucleus. Almost all of these materials can be converted by dehydrogenation with selenium into compounds which contain phenanthrene.

The molecular structure of a sterol (I), a steroid which contains a hydroxy or alcohol group (OH), is

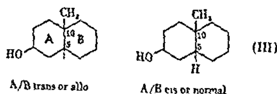


The abbreviated way of drawing the structure of this sterol, called dihydrocholesterol (II), is



The rings A, B, C, and D, which include carbon atoms numbered as shown from 1 through 17, are the cyclopentanoperhydrophenanthrene system. In carbon compounds, the carbon atom is at the center of a tetrahedral figure, and the four atoms or groups of atoms attached to the carbon are not in one plane. When all four atoms or groups are different, the compound can exist in two mirror-image forms, each of which is called an isomer. The carbon atom bearing four different substituents is termed an asymmetric carbon atom. Dihydrocholesterol contains nine centers of asymmetry, at carbon atoms 3, 5, 8, 9, 10, 13, 14, 17, and 20, and the number of possible isomers is 2^9, or 512 materials, each of which, though closely related, would differ from the next. Physiological activity is usually highly specific for a single isomer. Only a few of the total possible isomers for each compound with the same number of carbon, hydrogen, and ...

... terols are cholestanol and coprostanol, and these differ in respect to the configuration of rings A and B. In all steroids which contain the angular methyl groups (CH_3), carbon atoms 18 and 19, these project upward, or above the plane of the molecule. The configuration of an atom or group above the plane is shown as a solid line in the structural formula, and projection behind the plane of the molecule is shown as a dotted line. Rings A and B can be joined so that the hydrogen atom at carbon 5 projects to the rear, or behind the plane of the molecule as in cholestanol (III).



A/B trans or allo   A/B cis or normal

This trans or opposite spatial arrangement ...

... where the hydrogen atom at carbon atom 9 projects behind the plane, in all naturally occurring steroids. Similarly, rings C and D are joined in the trans configuration, with the hydrogen at carbon atom 14 projecting behind the plane, in all steroids except the cardiac active principles, which are in the cis arrangement, with the hydrogen at ...
the mole ...
carbon a ...
atoms 17 ...
occurrin ...

Source. Steroids are obtained from natural sources by thorough extraction with an organic solvent, usually ether. Separation of mixtures of steroids can be carried out by a variety of procedures, depending on the kind of steroid. These may involve the use of specific reagents like Girard's

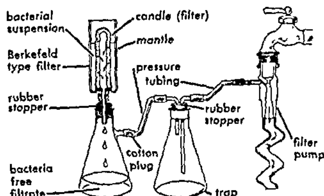


Fig. 4 Filtration sterilization by Berkefeld type of filter. (Redrawn from D. L. Belding and A. T. Marston, A Textbook of Medical Bacteriology, Appleton-Century-Crofts, 1938)

Filtration sterilization. This is the physical removal of microorganisms from liquids, by filtering through materials having relatively small pores. Sterilization by filtration is employed with liquid that may be destroyed by heat, such as blood serum, enzyme solutions, antibiotics, and some bacteriological media and medium constituents. Examples of such filters are the Berkefeld filter (diatomaceous earth), Pasteur-Chamberland filter (porcelain), Seitz filter (asbestos pad), and the sintered glass filter. Most of these filters are available in different pore sizes.

The mean pore size of bacteriological filters is not the only determinant in their effectiveness. The electric charge of the pore surfaces tends to adsorb the bacteria and thus prevent their passage. Most bacteria have a net negative electrical charge on their surfaces. Usually, bacteriological filters will permit the passage of viruses, which are then called filterable.

A Millipore filter is a specially prepared membrane molecular filter designed to remove bacteria from water, air, and other materials, for the purpose of estimating quantitatively the bacterial population. A sterile filter disk is assembled in a filtration unit and a specified volume of water or solution is drawn through the disk which then retains the bacteria. The filter disk is removed and placed in a sterile petri dish containing an absorbent pad, previously saturated with an appropriate bacteriological medium. Upon incubation, colonies will develop on the filter disk wherever bacteria were entrapped at the time of filtering. Special differential, or selective, media can be employed to detect quantitatively specific types of bacteria from the original material. See WATER ANALYSIS.

Chemical sterilization. Chemicals are used to sterilize solutions, air, or the surfaces of solids. Such chemicals are called bactericidal substances. In lower concentrations they become bacteriostatic rather than bactericidal, that is, they prevent the growth of bacteria but may not kill them. Other terms having similar meanings are employed. A disinfectant is a chemical that kills the vegetative cells of pathogenic microorganisms but not necessarily the endospores of sporeforming pathogens. An antiseptic is a chemical applied to living tissue that prevents or retards the growth of microorganisms, especially pathogenic bacteria, but which does not necessarily kill them.

The death of microorganisms subjected to bactericidal substances can be expressed exponentially, in that a straight-line graph is produced when the logarithm of survivors is plotted against time. The more concentrated the chemical employed, the greater is the rate of death.

There are hundreds of chemicals that may be considered to have sterilizing or bactericidal properties, depending upon the particular use for which they are intended. A chemical may be particularly useful for one purpose but not for another. Many are widely used for sterilization or disinfection of air, water, table tops, surgical instruments, and so on.

The desirable features sought in a chemical sterilizer are toxicity to microorganisms but nontoxicity to man and animals, stability, solubility, inability to react with extraneous organic materials, penetrative capacity, detergent capacity, noncorrosiveness, and minimal undesirable staining effects. Rarely does one chemical combine all these desirable features.

Among chemicals that have been found useful as sterilizing agents are the phenols, alcohols, chlorine compounds, iodine, heavy metals and metal complexes, dyes, and synthetic detergents, including the quaternary ammonium compounds.

Chlorine. Chlorine and chlorine-containing compounds represent the most widely used group of disinfectants. Chlorine gas is often used to purify municipal water.

Various compounds of chlorine, such as the hypochlorites and chloramine, have many industrial and domestic uses as disinfectants or antiseptics.

Ozone. This is a highly oxidizing gas (O_3) used as a deodorizer, and also for disinfection of air and water. It has found some use in the food fields, but effective bactericidal concentrations may be irritating and toxic to humans.

Hydrogen peroxide. This chemical (H_2O_2) has high oxidizing and bleaching qualities and is usually employed in a 3% solution for topical application and disinfection of cuts, scratches, and minor wounds. It decomposes into water and oxygen and therefore is used where no taste, odor, or toxic residues are permitted.

Volatile organic compounds. Such compounds as formaldehyde and ethylene oxide have been used for the disinfection of sickrooms occupied by patients suffering from contagious disease (terminal disinfection), and of solids that do not permit heat treatment. These volatile substances have the advantages of effective penetrative capacity and ease of removal, after treatment.

Volatile organic substances are sometimes employed as bacteriostatic agents to preserve bacteriological medium constituents until they are heat sterilized. An example is a mixture of chloro-

one or more fermentation steps. Some of the newer sex-hormone analogs sold as drugs are manufactured partly by fermentation. Thus, steroid fermentation products represent a large share of the \$120,000,000 estimated annual sales of steroid hormones in 1958 and 1959 [R.W.T.]

Bibliography: L. F. Fieser and M. Fieser, *Natural Products Related to Phenanthrene*, 3d ed., 1949; R. S Harris et al (eds), Synthetic derivatives of cortical hormones, *Vitamins and Hormones*, vol 16, 1958; G. Pincus (ed), The use of microorganisms in the synthesis of steroid hormones and hormone analogues, *Recent Progr. in Hormone Research*, vol. 11, 1955; C. W. Shoppee, *Chemistry of the Steroids*, 1958; L. A Underkofler and R. J. Hickey (eds), *Industrial Fermentations*, vol. 2, 1954; E Vischer and A Wettstein, Enzymic transformations of steroids by microorganisms, *Advances in Enzymol*, 20:237, 1958.

Sterol

One of the widely occurring steroids, which contains a hydroxyl or alcohol (OH) group at carbon atom 3 and an aliphatic side chain attached to carbon atom 17 of the cyclopentanoperhydrophenanthrene nucleus. The molecular structure of a sterol is shown in the formula. Specific examples of sterols are cholesterol, ergosterol, adrenal cortex steroids, bile acids, estrogens, androgens, and digitoxigenin

Dihydrocholesterol

These sterols may be classed as zoosterols (animal), phytosterols (plant), mycosterols (yeast and fungus), and marine sterols (sponge) Most animal sterols contain 27 carbon atoms, but lanosterol and agnosterol present in wool fat, are 30-carbon compounds *Plant and algal* sterols ordinarily have 29 carbon atoms, and yeast and fungal sterols are 27- or 28-carbon compounds No direct function is known for these compounds In animal tissues, relatively small amounts of cholesterol are used as the precursor, or starting material, for steroid hormones *See* ADRENAL CORTEX STEROID; ANDROGEN; ESTROGEN; PROGESTERONE.

Asymmetry in sterols. Asymmetry of the hydroxyl group at carbon atom 3 is possible A spatial arrangement in which this group extends above the plane of the molecule, on the same side as the angular methyl group, C-18, is termed the 3(β)

3(β)-Hydroxy. 3(α)-Hydroxy.

3(α) and 3(β) configurations

configuration When the hydroxyl group projects on the opposite side, this is known as the 3(α) configuration All natural sterols appear to be 3(β) compounds The substance [illegible] erty of forr [illegible] solution wit [illegible] tion of the [illegible] and analytical purposes.

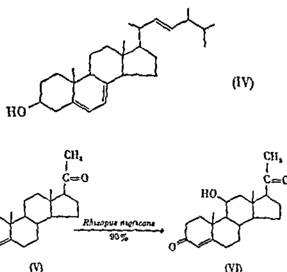
Sterol isolation. Sterols are isolated from natural sources by extraction with organic solvents, usually after treatment with alkali to saponify or produce free sterol from sterol esterified to long-chain fatty acids (*see* ESTERIFICATION). Crystallization of extracted sterols yields mixtures, separable with difficulty, of closely related materials, because molecular compounds and mixed crystals form readily. Chemical preparation of derivatives with differential solubilities, the use of chromatography, or other methods are necessary for adequate purification Analysis of sterols is based on the formation of colored materials by treatment with strong acids under dehydrating conditions. *See* CHROMATOGRAPHY.

Absorption and transport in blood. Dietary cholesterol is readily absorbed into the lymph system from the small intestine; however, only small amounts of plant sterols can be absorbed. About half of the total cholesterol in the lymph is in the esterified form. Cholesterol is emptied through the thoracic [illegible] conta[illegible] 2.) o[illegible] [illegible] for esterification of cholesterol. In the small intestine, an esterase catalyses the combination of cholesterol and a free fatty acid In the liver, a different enzyme joins a fatty acid, bound to coenzyme A, to cholesterol (*see* COENZYME). In the esterified cholesterol of blood plasma, almost all of the fatty acids are unsaturated (*see* UNSATURATED HYDROCARBON). Cholesterol and its ester are bound loosely and are transported in blood plasma, in association with lipoproteins The total cholesterol content in human plasma is about 200 mg per 100 ml. Subjects with higher levels show increased tendency toward coronary diseases

Excretion and transformation of sterols. The sterol mixture in animal feces consists of unabsorbed sterols ingested in the diet, coprostanol formed by bacterial reduction of cholesterol excreted from the bile into the small intestine, and traces of cholestanol. A large part of the loss of cholesterol from the body is by conversion to bile acids (*see* BILE ACID). This involves the removal of the end three carbon atoms of the side chain and modification of nuclear substituents Cholesterol is

reagent for ketonic (C=O) steroids, and digitonin for sterols, or by fractional crystallization, solvent partition, and chromatography. Identification of individual steroids is based on many criteria, including analysis for carbon and hydrogen, molecular-weight determination, melting point, optical activity, conversion by chemical methods to other known compounds, absorption spectroscopy in the ultraviolet and infrared portions of the spectrum, and by chemical synthesis. *See* ADRENAL CORTEX STEROID; ANDROGEN; BILE ACID; DIGITOXIGENIN; ESTROGEN. [I.Z.]

Industrial production by fermentation. In the fermentation processes for the production of steroids, certain biochemical activities of microorganisms are controlled so as to cause an abundant biosynthesis of a native sterol like ergosterol (IV) during growth or to modify some steroid such as progesterone (V) to a close analog 11-α-hydroxyprogesterone (VI) through the enzymatic action of the microorganism.

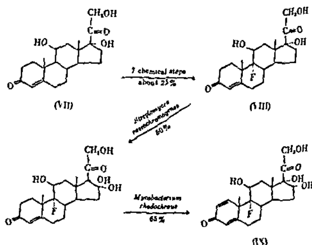


Ergosterol, the most important native microbial steroid, is a regular component of yeast (*Saccharomyces cerevisiae*), and thus may be recovered from the yeast crop during alcoholic fermentation as a by-product. Much higher yields, up to 2.7% of the yeast dry weight, can be obtained with certain strains grown in special, aerated media. Ergosterol may be recovered by digestion of the yeast cells with alkali, extraction of the resulting saponification mixture with a solvent such as ether, and crystallization from the ether. Ergosterol is marketed chiefly in the irradiated form, Viosterol, or vitamin D_2. The yearly production of irradiated ergosterol is about 1500 lb (25×10^{15} USP units), worth about $600,000. *See* ERGOSTEROL; VITAMIN D; YEAST, INDUSTRIAL.

Steroids not demonstrably native to microorganisms may be converted to more useful analogs by exposure to growing or pregrown cells, or to cell-free enzyme systems of selected species of microorganisms. Microorganisms useful in converting steroids are in general propagated in submerged culture under conditions similar to those employed in the production of antibiotics. Usually, the steroid substrate is added to the culture in a solvent such as methanol, propylene glycol, N,N-dimethyl formamide to give a concentration of steroid in the fermentation of 100–1000 mg/liter. Conversion efficiencies approaching the theoretical may be obtained in contact periods of a few hours to several days. Recovery of the conversion product or products is accomplished by solvent extraction of the whole or filtered fermentation mixture, concentration by evaporation, and recrystallization from a suitable nonaqueous solvent. If a mixture of products is obtained, purification by fractional crystallization or chromatography is required. *See* ANTIBIOTIC; CHROMATOGRAPHY; SOLVENT EXTRACTION.

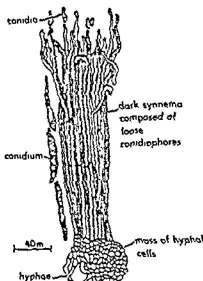
Microorganisms useful in the conversion of steroid substrates include a variety of yeasts, molds, bacteria, and actinomycetes. Most of these conversions involve oxidations or reductions, through elimination or addition of hydrogen or substitution of hydroxyl for hydrogen. The exact modification of the steroid depends upon the particular organism used and on the exposure conditions. *See* ACTINOMYCETALES; BACTERIA.

Several fermentation processes for conversion of steroids have been integrated into multistep syntheses of steroid hormones or hormone analogs. For example, the potent corticosteroid drug triamcinolone (IX), useful in the treatment of rheumatoid arthritis, may be produced from hydrocortisone (V) via fluorocortisone (VIII) in seven chemical steps followed by two fermentation steps, with an over-all yield of about 8% of theoretical. An alternative process with a different starting material, with chemical introduction of the 16-α hydroxyl group, might require four or more additional steps with less than one-half the over-all yield. Hydrocortisone itself may be a fermentation product; its synthesis via microbiological 11-α hydroxylation allows the use of relatively abundant plant steroids instead of animal steroids as raw materials. *See* ARTHRITIS; HORMONE.



Most of the corticosteroids, for example, cortisone, hydrocortisone, and newer analogs, are sold as antirheumatic, antiarthritic, anti-inflammatory drugs and are manufactured by processes involving

Isaria species are saprophytic or parasitic on insects. I. (*Paecilomyces*) *farinosa* is commonly found on dead insects *Coremium* is a genus which has synnemata with green, fertile heads and the conidia are in chains (*catenulate*). All species are doubtless stilboid stages of *Penicillium*. For example, *C. glaucum* belongs to *P. expansum*. *Graphium* has synnemata which are tall and dark, bearing a rounded, terminal mass of hyaline conidia embedded in mucilage. Many species are imperfect stages of *Ceratocystis* (*Ophiostoma*). *G. ulmi* causes Dutch elm disease



Isariopsis griseola Synnema composed of loose conidiophores. Conidia dark, two or more cells. (After G Viennot Bourgin, 1949)

The Phaeophragmiae have dark spores with two or more cells *Isariopsis* is a genus which has dark synnemata composed of loose conidiophores. There are 10 parasitic species known. *I griseola* causes a disease of beans *See* MONILIALES. [N.F.B]

Stilbite

A mineral belonging to the *zeolite family of silicates* It crystallizes in the monoclinic system in crystals that are tabular parallel to the side pinacoid Most characteristic are sheaflike aggregates of thin tabular crystals There is perfect cleavage parallel to the side pinacoid and here the mineral has a pearly luster, elsewhere the luster is vitreous The color is usually white but may be brown, red, or yellow Hardness is 3½–4 on Mohs scale; specific gravity is 2 1–2 2. *See* ZEOLITE.

Stilbite is a *calcium-sodium aluminum silicate*, $Ca(Al_2Si_7O_{18})\cdot 7H_2O$ The *ratio* of calcium to sodium varies and with it a corresponding variation in the amount of aluminum substituting for silicon Potassium is usually present substituting for sodium

Stilbite is a secondary mineral usually found in cavities in basalts and related rocks. Much less



Sheaflike aggregates of thin crystals typical of mineral stilbite (From C. S. Hurlbut, Jr., *Dana's Manual of Mineralogy*, 16th ed., Wiley, 1952)

commonly it is found in granites, gneisses, and in metal-bearing veins It is associated with other [illegible] New Jersey, and the Lake Superior copper district, Michigan. [C.S.HU.]

Stirodonta

An order of Euechinoidea proposed by R. Jackson in 1912. It included forms with keeled teeth and an open foramen magnum in the lantern. J Durham and R. Melville (1957) abandon the group as polyphyletic. *See* ARBACIOIDA; ECHINACEA; EUECHINOIDEA; HEMICIDAROIDA; PHYMOSOMATOIDA. [H.B F.]

Stochastic process

A physical stochastic process is any process governed by probabilistic laws. Examples are (1) development of a population as controlled by *Mendelian genetics*; (2) Brownian motion of microscopic particles subjected to molecular impacts, or, on a *different scale*, the motion of stars in space; (3) succession of plays in a gambling *house*; (4) passage of cars by a specified highway point

In each case, a probabilistic system is evolving, that is, its state is changing with time. Thus the state at time t depends on chance: it is a random variable $x(t)$. The parameter set of values of t involved is usually (and will always be in this article) either an interval (continuous parameter stochastic *process*) or a set of integers (*discrete* parameter stochastic process) Some authors, however, apply *the* term stochastic *process only to the* continuous parameter case

If the *state of the* system is described by a single number, $x(t)$ is numerical-valued In other *cases*, $x(t)$ *may* be vector-*valued or* even more complicated. The discussion in this article will usually be restricted to the numerical case As the state changes, its values determine a function of time, the sample function, and the probability laws governing the process determine the probabilities assigned to the various possible properties of sample functions.

A mathematical stochastic process is a mathematical structure inspired by *the* concept of a physical stochastic process, and studied because it is a mathematical model *of a physical* stochastic process, or because of its intrinsic mathematical interest and its applications both *in* and outside the field of probability. *The* mathematical stochastic process is defined simply as a family of random

converted in animal tissues to male and female sex hormones, progesterone, and adrenal hormones. Alternate paths from acetate to some of these hormones, without the intermediary formation of cholesterol, may also occur.

Pathology. Several pathological conditions, in which sterols have been implicated, are discussed in the following paragraphs.

In arteriosclerosis, the walls of arteries, particularly the intima, or innermost coat, show an abnormal thickening and hardening. Atherosclerosis is a type of arteriosclerosis in which the fibrous thickening of the intima is accompanied by atheromatous, or fatty, degeneration Atherosclerotic plaques contain lipids rich in cholesterol esters. This condition can readily be produced in rabbits by feeding diets rich in cholesterol. Atherosclerosis occurs more frequently in older humans, in whom serum cholesterol levels appear to be higher, than in younger persons.

Xanthomatoses are diseases characterized by the presence of multiple, benign fatty tumors, often rich in cholesterol, which are present in skin, tendon sheaths, and bone. In many cases, associated high levels of serum cholesterol are found. *See* CHOLESTEROL; DIGITOXIGENIN; STEROID [I.Z.]

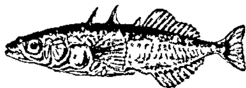
Stibnite

A mineral with composition Sb_2S_3 (antimony trisulfide), the chief ore of antimony. It crystallizes in the orthorhombic system, in slender, prismatic, vertically striated crystals which may be curved or bent. It is often in bladed, granular, or massive aggregates. There is one direction of perfect cleavage showing cross striations. The hardness is 2 and the specific gravity 4.5–4.6 The luster is metallic and the color lead-gray to black. It is one of few minerals that fuses easily in the match flame (525°C).

Stibnite is found in quartz veins in granite and gneiss with few other minerals present. Elsewhere it may be associated with cinnabar, realgar, orpiment, gold, galena, and sphalerite. Replacement deposits in limestone are probably the result of deposition by hot springs It has been found in various mining districts in Germany, Rumania, France, Bolivia, Peru, and Mexico Before World War II the most important commercial deposits were in the Province of Hunan, China. In the United States the Yellow Pine mine at Stibnite, Idaho, is the largest producer Other deposits are in Nevada and California. The finest crystals have come from the island of Shikoku, Japan *See* ANTIMONY. [C.S.HU.]

Stickleback

Any of about 12 species of small fishes of the family Gasterosteidae, found in both salt and fresh water, widely distributed in the Northern Hemisphere. The sticklebacks are characterized by a series of spines on the back, each with a small fin membrane, and by the very slender caudal peduncle They are known for their nest-building habit,



The three-spined stickleback, Gasterosteus aculeatus (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

the males building an oriolelike nest of aquatic plants held together by strands of mucus Each male guards his nest until the eggs have hatched

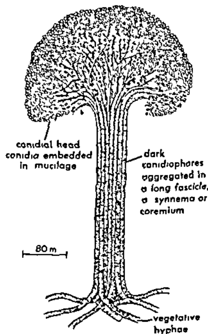
Sticklebacks live in shallow water and feed upon zooplankton and aquatic insects. In some localities they are of value in the control of mosquitoes *See* GASTEROSTEIFORMES. [J.D.B.]

Stilbellaceae

A family of fungi of the order Moniliales. Stilbaceae is a synonym for Stilbellaceae. The conidiophores are aggregated in long bundles or fascicles, forming so-called synnemata or coremia, generally having the conidia in a head at the top. The hyphae and conidia are hyaline or dark There are about 80 genera and 350 species known.

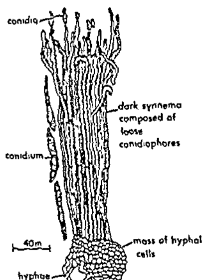
The genera are usually arranged into spore groups depending upon the characteristics of the spore, such as the number of cells in the spore, and the shape and pigmentation of the spore.

The Hyalosporae have 1-celled hyaline spores *Stilbella, Isaria, Coremium,* and *Graphium* are important genera in the Hyalosporae. *Stilbum* (*S. vulgare*) belongs to the Pilacraceae, a family of Basidiomycetes *Isaria* is a genus which has light-colored cylindrical or club-shaped (clavate) synnemata, forming conidia along the whole length



Graphium ulmi Synnema bearing a terminal mass of conidia embedded in mucilage. (*After C. Ferdinandsen and C. A. Jorgensen, 1938–1939*)

aria species are saprophytic or parasitic on insects *I* (*Paecilomyces*) *farinosa* is commonly found on dead insects *Coremium* is a genus which has synnemata with green, fertile heads and the conidia are in chains (catenulate). All species are doubtless stilboid stages of *Penicillium*. For example, *C glaucum* belongs to *P. expansum*. *Graphium* has synnemata which are tall and dark, bearing a rounded, terminal mass of hyaline conidia embedded in mucilage Many species are imperfect stages of *Ceratocystis* (*Ophiostoma*) *G. ulmi* causes Dutch elm disease



Isariopsis griseola Synnema composed of loose conidiophores, Conidia dark, two or more cells (After G Viennot Bourgin, 1949)

The Phae... or more cel... synnemata ... are 10 para... species known *I griseola* causes a disease of beans *See* MONILIALES [S F B]

Stilbite

A mineral belonging to the zeolite family of silicates It crystallizes in the monoclinic system in crystals that are tabular parallel to the side pinacoid Most characteristic are sheaflike aggregates of thin tabular crystals There is perfect cleavage parallel to the side pinacoid and here the ... has ... T ... ... on Mohs scale, specific gravity is 2 1–2 2 *See* ZEOLITE

Stilbite is a calcium sodium aluminum silicate, $Ca(Al_2Si_7O_{18})\cdot 7H_2O$ The ratio of calcium to sodium varies and with it a corresponding variation in the amount of aluminum substituting for silicon Potassium is usually present substituting for sodium

Stilbite is a secondary mineral usually found in cavities in basalts and related rocks Much less



Sheaflike aggregates of thin crystals typical of mineral stilbite. (From C. S. Hurlbut, Jr., Dana's Manual of Mineralogy, 16th ed, Wiley, 1952)

commonly it is found in granites, gneisses, and in metal bearing veins. It is associated with other ...

New Jersey, and the Lake Superior copper district, Michigan. [C.S.HU.]

Stirodonta

An order of Euechinoidea proposed by R. Jackson in 1912. It included forms with keeled teeth and an open foramen magnum in the lantern J. Durham and R. Melville (1957) abandon the group as polyphyletic. *See* ARBACIOIDA; ECHINACEA; EUECHINOIDEA, HEMICIDAROIDA; PHYMOSOMATOIDA [H B F.]

Stochastic process

A physical stochastic process is any process governed by probabilistic laws. Examples are (1) development of a population as controlled by Mendelian genetics; (2) Brownian motion of microscopic particles subjected to molecular impacts, or, on a different scale, the motion of stars in space; (3) succession of plays in a gambling house; (4) passage of cars by a specified highway point.

In each case, a probabilistic system is evolving, that is, its state is changing with time. Thus the state at time t depends on chance: it is a random variable $x(t)$ The parameter set of values of t involved is usually (and will always be in this article) either an interval (continuous parameter stochastic process) or a set of integers (discrete parameter stochastic process) Some authors, however, apply the term stochastic process only to the continuous parameter case.

If the state of the system is described by a single number, $x(t)$ is numerical valued In other cases, $x(t)$ may be vector-valued or even more complicated. The discussion in this article will usually be restricted to the numerical case As the state changes, its values determine a function of time, the sample function, and the probability laws governing the process determine the probabilities assigned to the various possible properties of sample functions.

A mathematical stochastic process is a mathematical structure inspired by the concept of a physical stochastic process, and studied because it is a mathematical model of a physical stochastic process, or because of its intrinsic mathematical interest and its applications both ...

converted in animal tissues to male and female sex hormones, progesterone, and adrenal hormones. Alternate paths from acetate to some of these hormones, without the intermediary formation of cholesterol, may also occur.

Pathology. Several pathological conditions, in which sterols have been implicated, are discussed in the following paragraphs.

In arteriosclerosis, the walls of arteries, particularly the intima, or innermost coat, show an abnormal thickening and hardening. Atherosclerosis is a type of arteriosclerosis in which the fibrous thickening of the intima is accompanied by atheromatous, or fatty, degeneration. Atherosclerotic plaques contain lipids rich in cholesterol esters. This condition can readily be produced in rabbits by feeding diets rich in cholesterol. Atherosclerosis occurs more frequently in older humans, in whom serum cholesterol levels appear to be higher, than in younger persons.

Xanthomatoses are diseases characterized by the presence of multiple, benign fatty tumors, often rich in cholesterol, which are present in skin, tendon sheaths, and bone. In many cases, associated high levels of serum cholesterol are found. *See* CHOLESTEROL, DIGITOXIGENIN; STEROID. [I Z.]

Stibnite

A mineral with composition Sb_2S_3 (antimony trisulfide), the chief ore of antimony. It crystallizes in the orthorhombic system, in slender, prismatic, vertically striated crystals which may be curved or bent. It is often in bladed, granular, or massive aggregates. There is one direction of perfect cleavage showing cross striations The hardness is 2 and the specific gravity 4 5–4.6. The luster is metallic and the color lead-gray to black It is one of few minerals that fuses easily in the match flame (525°C).

Stibnite is found in quartz veins in granite and gneiss with few other minerals present Elsewhere it may be associated with cinnabar, realgar, orpiment, gold, galena, and sphalerite. Replacement deposits in limestone are probably the result of deposition by hot springs It has been found in various mining districts in Germany, Rumania, France, Bolivia, Peru, and Mexico Before World War II the most important commercial deposits were in the Province of Hunan, China In the United States the Yellow Pine mine at Stibnite, Idaho, is the largest producer. Other deposits are in Nevada and California. The finest crystals have come from the island of Shikoku, Japan *See* ANTIMONY. [C S HU]

Stickleback

Any of about 12 species of small fishes of the family Gasterosteidae, found in both salt and fresh water, widely distributed in the Northern Hemisphere. The sticklebacks are characterized by a series of spines on the back, each with a small fin membrane, and by the very slender caudal peduncle. They are known for their nest-building habit,



The three-spined stickleback, Gasterosteus aculeatus (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

the males building an oriolelike nest of aquatic plants held together by strands of mucus. Each male guards his nest until the eggs have hatched.

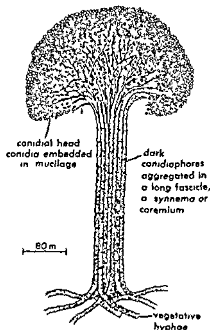
Sticklebacks live in shallow water and feed upon zooplankton and aquatic insects. In some localities, they are of value in the control of mosquitoes. *See* GASTEROSTEIFORMES. [J D B]

Stilbellaceae

A family of fungi of the order Moniliales. Stilbaceae is a synonym for Stilbellaceae The conidiophores are aggregated in long bundles or fascicles, forming so-called synnemata or coremia, generally having the conidia in a head at the top. The hyphae and conidia are hyaline or dark. There are about 80 genera and 350 species known.

The genera are usually arranged into spore groups depending upon the characteristics of the spore, such as the number of cells in the spore, and the shape and pigmentation of the spore.

The Hyalosporae have 1-celled hyaline spores *Stilbella*, *Isaria*, *Coremium*, and *Graphium* are important genera in the Hyalosporae. *Stilbum* (*S. vulgare*) belongs to the Pilacraceae, a family of Basidiomycetes. *Isaria* is a genus which has light-colored cylindrical or club-shaped (clavate) synnemata, forming conidia along the whole length



Graphium ulmi. Synnema bearing a terminal mass of conidia embedded in mucilage (After C Ferdinandsen and C. A. Jorgensen, 1938–1939)

eter ranging from 0 to ∞ can be constructed for which, if again p_j is the probability that $x(0)$ has the value j, and if $0 < t_1 < \cdots < t_N$, the probability that $x(t_k)$ has the value i_k for $k = 1, \ldots, N$ is given by

$$\sum_i p_i p_{ii_1}(t_1) p_{i_1 i_2}(t_2 - t_1) \cdots p_{i_{N-1} i_N}(t_N - t_{N-1}) \quad (5)$$

For this process, if $s > 0$, the probability that $x(u+s)$ has the value j, if $x(u)$ has the value i, is $p_{ij}(s)$. The number $p_{ij}(s)$ is also the probability (and this is the characteristic property of Markov processes) that $x(u+s)$ has the value j if $x(u)$ has the value i, and if also $x(u_1)$ has any specified value a_1, $x(u_2)$ the value a_2, and so on, where $u_1, u_2, \ldots$ are any positive numbers less than u. This example is not the general continuous-parameter Markov chain because the transition probability just described does not depend on u, that is, because the chain has stationary transition probabilities The process is stationary if the probability that $x(u)$ has the value j does not depend on u, for all j The second line of Eq (4) has a simple interpretation: the probability, if $x(u)$ has the value i, that $x(u+s+t)$ has the value j, is the sum over k of the probability that $x(u+s)$ has the value k multiplied by the probability that, if $x(u)$ has the value i, and if $x(u+s)$ has the value k, then $x(u+s+t)$ has the value j Without the Markov property, the second factor might depend on i. If the number of states is not finite or denumerably infinite, the preceding discussion is modified by replacing sums in Eqs (1), (3), (4), and (5) by integrals

Typical applications Typical questions that have been raised, and solved to a varying degree, about Markov processes with stationary transition probabilities, are the following They are phrased in the continuous parameter case, for the Markov chain just described, and it is assumed that

$$\lim_{t \to 0} p_{ii}(t) = 1 \quad \text{for all } i$$

For convenience one defines $p_{ij}(0)$ as 1 if $i = j$ and as 0 otherwise

1 Does $p_{ij}(t)$ have a limit when $t \to \infty$? In other words, for each j is there a limiting probability that the system is in state j as time passes? The answer is yes, and the limiting probability depends only on the end state j, not on the initial state i, if transitions between all pairs of states are possible.

2 What are the asymptotic properties of $p_{ij}(t)$ as $t \to 0$? The answer is that $p'_{ij}(0)$ always exists, and is finite except possibly when $i = j$. Under further hypotheses on the transition probability functions, always satisfied if there are only finitely many states, $p'_{ii}(0)$ is finite, and

$$p'_{ij}(t) = \sum_k p_{ik}(t) p'_{kj}(0)$$
$$p'_{ij}(t) = \sum_k p'_{ik}(0) p_{kj}(t) \quad (6)$$

These equations can be used to determine the transition probability functions in terms of assigned derivatives when $t = 0$. For example, if c is a strictly positive constant, and if $p'_{ij}(0)$ is specified as c if $j = i+1$, as $-c$ if $j = i$, and as 0 otherwise, it is shown that $p_{ij}(t) = 0$ if $j < i$, and that otherwise

$$p_{ij}(t) = \frac{(ct)^{j-i} e^{-ct}}{(j-i)!} \quad (7)$$

The process with these transition probabilities is known as the Poisson process.

3. What are the properties of the sample functions? Under further restrictions on the process, the sample functions are constant on intervals, changing in jumps from one state to the next, and $-p'_{ij}(0)/p'_{ii}(0)$ is the probability that, if the system is in state i, its next jump will be into state j If it is in the ith state, the time the system remains in this state thereafter is a random variable with density qe^{-tq}, where $q = -p'_{ii}(0)$. For example, in the case of the Poisson process described above, it is shown that, under proper normalization, and if $x(0) = 0$, the sample functions are integral-valued and monotone, increasing in unit jumps. This process is a mathematical model for the physical process of radioactive decay. That is, $x(t)$ can be interpreted as the number of radioactive disintegrations of a substance by time t. In other interpretations, $x(t)$ is taken as the number of telephone calls initiated by time t, or the number of cars that have passed a given highway point by time t. The constant c is the rate at which these various events occur In fact the expected value of $x(u+h) - x(u)$, that is, the expected number of events in a time interval of length h (h here is of course positive), is ch, and the probability that an event will occur in an interval of length h, regardless of the past history of the process, is ch up to higher powers of h.

There are many special types of Markov chains, for which more detailed questions become important. For example, consider the branching processes. In a system of particles, all of the same type, a particle will occasionally split, independently of its past history and of the other particles, into $j \geqq 0$ particles with probability q_j. If a particle is observed at time t, the probability that it will split by time $t+h$ is ch, up to higher powers of h, where c is a strictly positive constant. Then the number of particles at time t is a random variable $x(t)$. The $x(t)$ process is a Markov process, and $p'_{ij}(0)$ is easily determined in terms of q_j and c. A more general branching process would permit particles of several types, and each particle would be allowed to split into particles of the various types The rate c would depend on the particle type. In this case, $x(t)$ is defined as a vector, whose ith component is the number of particles of type i at time t. The $x(t)$ stochastic process is a vector-valued Markov process In studying branching processes, the most natural questions to ask are: What is the probability that the population will die

variables. That is, a parameter set is specified, and to each parameter point t a random variable $x(t)$ is specified. If one recalls that a random variable is itself a function, if one denotes a point of the domain of the random variable $x(t)$ by ω, and if one denotes the value of this random variable at ω by $x(t,\omega)$, it results that the stochastic process is completely specified by the function of the pair (t,ω) just defined, together with the assignment of probabilities. If t is fixed, this function of two variables defines a function of ω, namely the random variable denoted by $x(t)$. If ω is fixed, this function of two variables defines a function of t, a sample function of the process.

Probabilities are ordinarily assigned to a stochastic process by assigning joint probability distributions to its random variables These joint distributions, together with the probabilities derived from them, can be interpreted as probabilities of properties of sample functions. For example, if t_0 *is a parameter value, the probability* that a sample function is positive at time t_0 is the probability that the random variable $x(t_0)$ has a positive value The fundamental theorem at this level is that, to any self-consistent assignment of joint probability distributions, there corresponds a stochastic process.

The concept of a stochastic process is so general that the study of stochastic processes includes all of probability theory Although it would be impossible to make specifically a more restrictive definition, what the probabilist usually has in mind in using the term stochastic process (unless he is interested in the mathematical foundations of the general theory) is a stochastic process whose random variables have some sort of interesting mutual relations. For example, one such relation is that of independence, and one type of stochastic process which was studied long before the term was invented is a sequence of independent random variables. For historical reasons, such a sequence is not commonly thought of as a stochastic process (although its *continuous parameter analog, a process* with independent increments, to be discussed below, is). The rest of this article is devoted to a general discussion of specific types of stochastic processes which have received the most attention, because they are important in mathematical and nonmathematical applications.

Stationary processes. These are the stochastic processes for which the joint distribution of any finite number of the random variables is unaffected by translations of the parameter; that is, the distribution of $x(t_1+h), \ldots, x(t_n+h)$ does not depend on h. For a more complete discussion, *see* PROBABILITY.

Markov processes. A Markov process is a process for which, if the present *is given, the future and* past are independent of each other. More precisely, if $t_1 < \cdots < t_n$ are parameter values, and if $1 < j < n$, then the sets of random variables

$$[x(t_1), \ldots, x(t_{j-1})] \quad \text{and} \quad [x(t_{j+1}), \ldots, x(t_n)]$$

are mutually independent for given $x(t_j)$. Equivalently, the conditional probability distribution of $x(t_n)$ for given $x(t_1), \ldots, x(t_{n-1})$ depends only on the specified value of $x(t_{n-1})$, and is in fact the conditional probability distribution of $x(t_n)$, given $x(t_{n-1})$. An important and simple example is the Markov chain, in which the number of states is finite or denumerably infinite. (The terminology varies somewhat here) One simple type of discrete-parameter Markov chain is the following Let (p_{ij}) be a set of numbers, where i, j range over a finite or infinite set of integers. (In physical language, the number p_{ij} will be the probability that some system has a transition from state i to state j in one step) The numbers p_{ij} are to satisfy

$$p_{ij} \geq 0 \qquad \sum_j p_{ij} = 1 \tag{1}$$

The random variables of the associated Markov process are integral-valued, denoted by $x(0)$, $x(1), \ldots$. If i_0 is prescribed as the *initial state*, that is, if $x(0)$ is assigned the value i_0 identically, the probability that $x(k)$ has the value i_k, for $k = 1, \ldots, N$, is the product

$$p_{i_0 i_1} \cdots p_{i_{N-1} i_N} \tag{2}$$

For this process, p_{ij} is the probability that $x(n+1)$ has the value j if $x(n)$ has the value i The number p_{ij} is also (and this is the characteristic property of Markov processes) the probability that $x(n+1)$ has the value j if $x(n)$ has the value i and if also $x(n-1)$ has any prescribed value a_1, $x(n-2)$ the value a_2, and so on. What makes this a special Markov chain, aside from the fact that the states are denoted by integers *is that* the conditional probability just described does not depend on n. The chain is therefore described as having stationary transition probabilities. If the initial state is given a distribution, say by prescribing that $x(0)$ have the value i with probability p_i, the evaluation (2) becomes

$$\sum_i p_i p_{i i_1} \cdots p_{i_{N-1} i_N} \tag{3}$$

Here the p_i's are any nonnegative numbers with sum 1 If the initial distribution is chosen, as is not always possible, in such a way that the probability that $x(n)$ has the value j is p_j, not only for $n = 0$ but for all values of n, the resulting process is stationary.

In constructing the corresponding continuous-parameter Markov chain, it is supposed that, for each pair (i,j), there is a function $p_{ij}(.)$, defined for strictly positive t, satisfying

$$\begin{gathered} p_{ij}(t) \geq 0 \qquad \sum_j p_{ij}(t) = 1 \\ p_{ij}(s+t) = \sum_k p_{ik}(s) p_{kj}(t) \end{gathered} \tag{4}$$

The equations of the system in the last line are known as the Chapman-Kolmogorov equations. A Markov stochastic process with continuous param-

cludes heat of reaction, heat added, heat lost, and other energy effects associated with the reaction. The third group comprises the equilibrium relationships of the system, and includes not only chemical equilibria but also physical equilibria such as gas-liquid systems. The fourth group includes the rate of reaction relationships of both chemical changes and physical changes.

In the chemical laboratory only the first group is usually considered. The second and fourth groups are neglected, or the reactions are forced to completion artificially. The third group is utilized in a limited number of cases. The units used are grams, moles, milliliters, or liters. The types of calculations include weight-weight, weight-volume (either gases or liquids), and thermal relationships. For any calculation a knowledge of the reaction under consideration and a balanced reaction equation are necessary.

Weight-weight problems. In these, a weight of material is given and a weight of product is desired. These calculations are based on the mole ratios shown in the balanced reaction equation.

Example 1. How many grams of calcium carbonate can be prepared from 100 grams (g) of sodium carbonate?

$$CaCl_2 + Na_2CO_3 = CaCO_3 + 2NaCl$$

The equation indicates that 1 mole of sodium carbonate will yield 1 mole of calcium carbonate.

$$\frac{100 \text{ g } Na_2CO_3}{\text{mole wt } Na_2CO_3} = \frac{x \text{ g } CaCO_3}{\text{mole wt } CaCO_3}$$

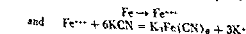
$$x \text{ g } CaCO_3 = 100 \text{ g } Na_2CO_3 \times \frac{\text{mole wt } CaCO_3}{\text{mole wt } Na_2CO_3}$$



$$x \text{ g} = 100 \times \frac{100}{106} = 94.3 \text{ g}$$

In this example the desired weight is equal to the given weight multiplied by a definite number, the ratio of molecular weights. This definite number is called a factor or gravimetric factor and is defined as the ratio of the molecular weight of the substance sought divided by the molecular weight of the substance weighed in the correct proportion. The correct proportion is found from a balanced reaction equation by means of an atom which is common to both molecules or by following a series of reactions.

Example 2: How many grams of potassium cyanide are required to complex the ferric ion obtained from 5 g of iron as potassium ferricyanide?



$$Fe \rightarrow Fe^{+++}$$

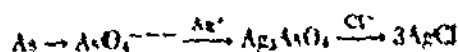
and $$Fe^{+++} + 6KCN = K_3Fe(CN)_6 + 3K^+$$

The equation says that for each atomic weight (at. wt) of iron, 6 moles of potassium cyanide is required.

$$\text{g KCN} = \text{g Fe} \times \frac{6 \text{ mole wt KCN}}{\text{at. wt Fe}}$$

$$= 5 \times \frac{6 \times 65.1}{55.85} = 34.97 \text{ g}$$

Example 3: A method for the isolation and determination of arsenic is based on oxidation to arsenate, precipitation of silver arsenate, and conversion of this to silver chloride which is weighed. If 1.325 g of silver chloride is found, what weight of arsenic was present?



$$As \rightarrow AsO_4^{---} \xrightarrow{Ag^+} Ag_3AsO_4 \xrightarrow{Cl^-} 3AgCl$$

In this case no common ion is present. However, arsenic which is sought is present in silver arsenate with silver which is weighed. Since the Ag to As ratio is 3:1 in silver arsenate, 3 moles of silver chloride would be weighed for each gram atom (g-atom) of arsenic present at the beginning.

$$\text{g As} = \text{g AgCl} \times \frac{\text{at wt As}}{3 \text{ mole wt AgCl}}$$

$$= 1.325 \times \frac{74.91}{3 \times 143.4} = 0.2307 \text{ g}$$

Weight-volume calculations. These fall into two types according to whether the volume deals with a gas or a solution. Problems dealing with gases are best solved in terms of molar volumes, the volume occupied by 1 mole of gas at 0°C and 760 millimeters pressure (standard conditions, STP). This volume is 22.4 liters. The calculated volumes can be corrected to any stated conditions by use of the gas laws. Gravimetric factors are used as in weight-weight problems except that molar volume will replace molecular weight for one substance.

Example 4: How many liters of dry HCl gas at STP is needed to convert 10 g of ethylamine to ethylamine hydrochloride?



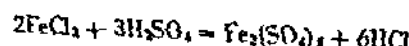
$$C_2H_5NH_2 + HCl = C_2H_5NH_2Cl$$

$$x \text{ liters HCl} = \text{g } C_2H_5NH_2 \times \frac{\text{molar vol HCl}}{\text{mole wt } C_2H_5NH_2}$$

$$= 10 \times \frac{22.4}{45.0} = 5.0 \text{ liters}$$

Calculations for reactions involving solutions are worked in terms of moles also. Molarity, M, is defined as the number of moles of solute per liter of solution. Molarity can also be calculated from the density, grams of solute per milliliter, and the volume in milliliters.

Example 5. How many liters of sulfuric acid, density 1.84 and 96% H_2SO_4 by weight, is needed to convert 75.6 g of ferric chloride to ferric sulfate?



$$2FeCl_3 + 3H_2SO_4 = Fe_2(SO_4)_3 + 6HCl$$

$$x \text{ g } H_2SO_4 = \text{g } FeCl_3 \times \frac{3 \text{ mole wt } H_2SO_4}{2 \text{ mole wt } FeCl_3}$$

$$V_{\text{liter}} H_2SO_4 = x \text{ g } H_2SO_4$$

$$\times \frac{1}{\text{density(g/ml)} \times (\text{fraction } H_2SO_4)} \frac{1}{1000 \text{ ml/liter}}$$

$$V_{\text{liter}} = 75.6 \times \frac{3 \times 98}{2 \times 163} \times \frac{1}{1.84 \times 0.96} \times \frac{1}{1000}$$

$$= 0.0386 \text{ liters}$$

out? If it does not die out, what is the asymptotic distribution of population as time passes? The answers are too technical to be given here.

Transition probabilities. If the states of a Markov process comprise all real numbers, the character of the process may be similar to that of a chain but may also be quite different For example, the sample functions of the process may be continuous. The most important examples of this type are the diffusion processes. Simplifying somewhat, but not assuming stationary transition probabilities, consider a Markov process for which the probability distribution of the state at time t, given state ξ at time $s < t$, has density $p(s,\xi,t,.)$. Then the basic conditions satisfied by the transition density, corresponding to Eq. (1), are

$$p(s,\xi,t,\eta) \geq 0 \qquad \int_{-\infty}^{\infty} p(s,\xi,t,\eta)\, d\eta = 1 \tag{8}$$

$$p(s,\xi,t,\eta) = \int_{-\infty}^{\infty} p(s,\xi,u,\xi)p(u,\xi,t,\eta)\, d\xi \quad \text{if } s < t < u$$

Now suppose that the following limits exist and have the indicated values:

$$\lim_{h \downarrow 0} \int_{-\infty}^{\infty} p(s,\ \xi,\ s+h,\ \eta)(\eta - \xi)\, d\eta/h = m(s,\xi) \tag{9}$$

$$\lim_{h \downarrow 0} \int_{-\infty}^{\infty} p(s,\ \xi,\ s+h,\ \eta)(\eta - \xi)^2\, d\eta/h = \sigma^2(s,\xi) \tag{10}$$

These limit relations make m and σ^2 the instantaneous rates of change of the displacement and its variance, given a specified time and state. If m and σ are sufficiently regular, and if a further condition is imposed which, roughly, makes improbable significant sample function changes in short times, the corresponding Markov process, when properly normalized, will have continuous sample functions Moreover the transition density will then satisfy the backward-diffusion equation

$$\frac{\partial p(s,\xi,t,\eta)}{\partial s} + m(s,\xi)\frac{\partial p}{\partial \xi} + \frac{1}{2}\sigma^2(s,\xi)\frac{\partial^2 p}{\partial \xi^2} = 0 \tag{11}$$

and the forward-diffusion equation, also known as the Fokker-Planck equation,

$$\frac{\partial p(s,\xi,t,\eta)}{\partial t} + \frac{\partial}{\partial \eta}[m(t,\eta)p] - \frac{1}{2}\frac{\partial^2}{\partial \eta^2}[\sigma^2(t,\eta)p] = 0 \tag{12}$$

Conversely, given a pair of coefficient functions m and σ, these second-order parabolic equations can be used to derive the corresponding transition densities.

The simplest nontrivial example of a diffusion process corresponds to the specification $m = 0$ and σ a constant function. In this case, the diffusion process is the Brownian motion process, or Wiener process: the increment $x(t) - x(s)$ has a Gaussian distribution with mean value 0 and variance $\sigma^2|t - s|$. This is a mathematical model for the physical Brownian motion. That is, if $x(t) - x(s)$ represents the displacement in a given direction of a Brownian particle between times s and t, this process is a good model for the actual motion.

Martingales. A martingale is a stochastic process with the property that, if $t_1 < \cdots < t_n$ are parameter values, the expected value of $x(t_n)$, for given $x(t_1), \ldots, x(t_{n-1})$ is equal to $x(t_{n-1})$. That is, the expected future value, given present and past values, is equal to the present value. The interpretation that a martingale can be thought of as the fortune of a player after the successive plays of a fair gambling game is obvious.

Typical results on martingales are the following If a sequence of random variables is a martingale, it converges under weak conditions which impose certain "bounds" on the random variables, for example, if the expectation of the absolute value of the nth random variable is bounded independently of n The sample functions of a properly normalized continuous-parameter martingale do not have oscillatory discontinuities.

The applications of martingale theory are too technical to be given here, but one suggestive example will be given, which indicates at least how the theory can be usefully applied to information theory. Let $y, x_1, x_2, \ldots$ be random variables, and let y_n be the expected value of y knowing $x_1, \ldots, x_n$. Then y_n is a random variable which is a function of $x_1, \ldots, x_n$ and the sequence $y_1, y_2, \ldots$ is a martingale That is, the expected value of a random variable, if one knows more and more, defines a sequence of random variables which *is a martingale.*

Processes with independent increments. Such a process is a continuous-parameter process with the property that, if $t_1 < \cdots < t_n$ are parameter values, the successive increments

$$x(t_2) - x(t_1),\ \ldots,\ x(t_n) - x(t_{n-1})$$

are mutually independent. If $y(t) = x(t) - x(t_0)$, where t_0 is fixed, the $y(t)$ process is then a Markov process Both the Poisson and the Brownian motion *processes described above have independent increments*

Typical results on these processes include the following. If such a process is properly normalized its sample functions do not have oscillatory discontinuities Moreover the distribution of any increment $x(t) - x(s)$ is infinitely divisible, and there is a standard form for the characteristic function of any such distribution. *See* GAME THEORY; INFORMATION THEORY, LINEAR PROGRAMMING; OPERATIONS RESEARCH [J.L.D]

Bibliography: M S Bartlett, *An Introduction to Stochastic Processes*, 1955; J L. Doob, *Stochastic Processes*, 1953.

Stoichiometry

The interpretation of mass and energy relationships indicated by chemical reaction equations It is also defined as those calculations used to find the quantities of reactants and products in a chemical reaction

Principles. These fall into four groups. The first group includes the law of conservation of matter, the law of chemical combining weights, and the law of combining proportions. The second group is based on the law of conservation of energy and in

For precipitation and complex-formation titrations the equivalent weight is the number of grams which correspond to a single charge on the metal ion involved. In practice it is often the same as acid base equivalent weight.

Heat calculations. Occasionally the question of heat must be considered in laboratory work. Heat quantities are usually expressed in units of *calories* or *kilocalories*. Heat added, heat of reaction, heat of cooling, and heat uptake by both reactants and [illegible]

principles must be used in *industrial stoichiometry*. Flowing systems require the use of reaction rates and detailed considerations of heat losses, purity of materials, effects of catalysts, and side reactions. For example, the complete analysis of a plant engaged in the manufacture of sulfuric acid, starting with iron pyrites as the source of sulfur, is a time-consuming task. However, modern chemical industry is based on these detailed considerations. See CONCENTRATION SCALES; MOLAR VOLUME; OXIDATION REDUCTION; THERMOCHEMISTRY.

[K C S]

Bibliography: L. F. Hamilton and S. G. Simpson, *Calculations of Analytical Chemistry*, 5th ed., 1954; W. K. Lewis, A. H. Radasch, and H. C. Lewis, *Industrial Stoichiometry*, 2d ed., 1954.

Stoker

Mechanical means for handling coal into a furnace. Stokers are of three basic types. In moving grate stokers the grate on which the coal burns carries the coal by continuous or reciprocating motion from a hopper into the furnace. Such stokers also move the ash out of the furnace. In spreader stokers the coal is mechanically or pneumatically distributed from the front furnace wall onto the grate. The grate usually moves to dispose of the ash after combustion. In underfeed stokers used in small furnaces, the stoker, usually a screw conveyor, forces fresh coal up under the burning coal; the ash is, in turn, forced off peripherally to the ash pit around the edge of the retort, or removed by hand. See FURNACE (STEAM GENERATING). [R M.H.]

Stokes stream function

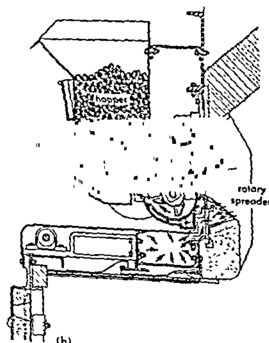
A degenerate (one-component) vector potential used in analyzing and describing axially symmetric fluid-flow fields. In a steady axially symmetric flow, the rotation of a streamline about the axis of symmetry generates a stream surface. A certain mass rate of flow exists inside this stream surface which is the same at every axial station, because, by definition, there is no flow through the stream surface. The value of the Stokes stream function ψ at a point in the flow is equal to $1/2\pi$ times mass rate of flow inside the stream surface passing through that point. If r is the radial coordinate of a point, and x the axial coordinate, then $\psi = \psi(x,r) =$ constant is the equation of a stream surface [illegible]

[illegible] them and of density ρ is

$$2\pi\, d\psi = 2\pi \frac{\partial \psi}{\partial r}\, dr = (2\pi r\, dr)\rho u,$$

hopper

(a)



Principal types of stokers. (a) Moving grate. (b) Spreader. (c) Underfeed

Normality. In quantitative analysis, volume calculations are cumbersome by the mole method. In order that equal volumes of equal concentration be chemically equivalent, a system based on equivalents is used. Normality, N, is defined as the number of equivalents (equiv) of solute per liter of solution. From this definition the following relationships are derived.

$$N = \frac{\text{No. of equiv}}{\text{liter}} = \frac{\text{No. of milliequivalents (meq)}}{\text{ml}}$$

$$\text{No. of equiv} = N \times V_{\text{liter}}$$

$$= \frac{\text{grams}}{\text{equiv wt}} \quad \frac{V_{\text{ml}} \times \text{density} \times \text{fraction}}{\text{equiv wt}}$$

$$\text{grams} = N \times V_{\text{liter}} \times \text{equiv wt} = N \times V_{\text{ml}} \times \text{meq wt}$$

These relationships are used to solve most volume calculations. The only quantity which varies according to the particular reaction is the equivalent weight. For acids, bases, and salts which are acidic or basic the equivalent weight is the number of grams which contain or react with 1 g-atom of replaceable hydrogen.

Example 6: What are the equivalent weights for example 4?

HCl has one replaceable hydrogen ion, its mole wt is 35.45; $C_2H_5NH_2$ combines with one H^+, its mole wt is 45.07 g.

Example 7: How many grams of NaOH is required to prepare 5 liters of 0.3 N solution? How many milliliters of 0.06 N H_2SO_4 will be neutralized by 75 ml of the NaOH solution?

NaOH reacts with one H^+, and its equiv wt is equal to its mole wt.

5 liters × 0.3 equiv/liter × 40 g/equiv = 60 g NaOH needed;

$$V_{\text{ml}} \times 0.06\ N = \text{no of meq } H_2SO_4$$

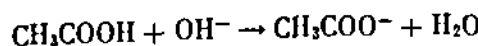
$$= \text{no of meq NaOH} = 75 \times 0.3$$

$$V_{\text{ml}} = \frac{75 \times 0.3}{0.06} = 375 \text{ ml } H_2SO_4 \text{ solution}$$

Example 8: If a 2.000-g sample of dilute acetic acid requires 40.00 ml of 0.2250 N base for neutralization, what per cent acetic acid is present?

$$\%\ \text{acetic acid} = \frac{\text{g acetic acid}}{\text{g sample}} \times 100$$

$$= \frac{N \times V_{\text{ml}} \times \text{meq wt} \times 100}{\text{g sample}}$$



$$CH_3COOH + OH^- \rightarrow CH_3COO^- + H_2O$$

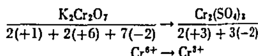
so that equiv wt of acetic acid equals mole wt.

$$\%\ \text{acetic acid} = 40.00 \times 0.2250 \times \frac{60.05}{1000} \times \frac{100}{2}$$

$$= 27.02\%$$

For oxidation-reduction systems, the equivalent weight is found by dividing the mole wt of a substance by its change in oxidation number during a reaction. Oxidation number is the apparent oxidation state or apparent valence of the atom The rules for calculating oxidation number may be summarized as follows. Elements have oxidation numbers of zero; single valent elements have an oxidation number of +1; oxygen is −2, except in peroxides in which it is −1; the sum of oxidation numbers for a compound is zero; and two atoms of the same element attached together have a zero contribution. For organic compounds, only the atoms which change need to be considered

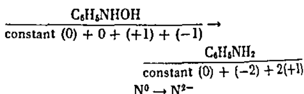
Example 9:



$$\frac{K_2Cr_2O_7}{2(+1) + 2(+6) + 7(-2)} \rightarrow \frac{Cr_2(SO_4)_3}{2(+3) + 3(-2)}$$

$$Cr^{6+} \rightarrow Cr^{3+}$$

The oxidation state of chromium changes by 3 *units per atom, $K_2Cr_2O_7$ has 2 Cr atoms;* for $K_2Cr_2O_7$ the equiv wt is 1/6 the mole wt.



$$\frac{C_6H_5NHOH}{\text{constant } (0) + 0 + (+1) + (-1)} \rightarrow \frac{C_6H_5NH_2}{\text{constant } (0) + (-2) + 2(+1)}$$

$$N^0 \rightarrow N^{2-}$$

The oxidation state of nitrogen changes by 2 units per atom, equiv wt for C_6H_5NHOH is 1/2 mole wt

Example 10: How many grams of $KMnO_4$ is present in 1 liter of 0.1 N solution to be used in acid medium? How many grams of ferrous ion will be oxidized by this solution?

In acid $KMnO_4 \rightarrow Mn^{++}$, $Mn^{7+} \rightarrow Mn^{2+}$ the oxidation state of manganese changes by 5 units per atom

$$\text{g } KMnO_4 = N \times V_{\text{liter}} \times \text{equiv wt}$$

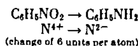
$$= 0.1 \times 1 \times 31.61 = 3.161$$

Fe^{++} is oxidized to Fe^{+++}, change of 1 in oxidation state.

$$\text{g } Fe^{++} = N \times V_{\text{liter}} \times \text{equiv wt}$$

$$= 0.1 \times 1 \times 55.85 = 5.585$$

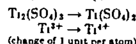
Example 13: What weight of $Ti_2(SO_4)_3$ is required to reduce 25 g of nitrobenzene to aniline?



$$C_6H_5NO_2 \rightarrow C_6H_5NH_2$$

$$N^{4+} \rightarrow N^{2-}$$

(change of 6 units per atom)



$$Ti_2(SO_4)_3 \rightarrow Ti(SO_4)_2$$

$$Ti^{3+} \rightarrow Ti^{4+}$$

(change of 1 unit per atom)

$$\text{No. of equiv } C_6H_5NO_2 = \text{No of equiv } Ti_2(SO_4)_3$$

$$\frac{\dfrac{\text{g } C_6H_5NO_2}{\text{mole wt } C_6H_5NO_2}}{6} = \frac{\dfrac{\text{g } Ti_2(SO_4)_3}{\text{mole wt } Ti_2(SO_4)_3}}{2}$$

$$\text{g } Ti_2(SO_4)_3 = \text{g } C_6H_5NO_2 \times \frac{6}{2} \times \frac{\text{mole wt } Ti_2(SO_4)_3}{\text{mole wt } C_6H_5NO_2}$$

$$= 25 \times \frac{6}{2} \times \frac{383.8}{123.1} = 233.8$$

fusion of the spicules. These tubes are united by *transverse* stolons or solid platforms at several levels See ALCYONARIA [K.A.]

Stomach

The tubular or saccular abdominal organ of the digestive system adapted for temporary food storage and preliminary stages of food breakdown

In lower vertebrates, such as fishes and amphibians, the stomach is little more than a simple tube quite similar to other portions of the gastrointestinal tract Small pockets, or diverticula, first appear in the reptiles *See* GASTROINTESTINAL TRACT.

In birds the stomach consists of a proventriculus and a gizzard The former is well supplied with glands which secrete softening and digestive materials, the latter is a strong muscular grinding organ whose action is often enhanced by the inges-

though a single chamber is most common, some mammals, such as the ungulates, have as many as four These may have developed either from modifications of the lower portions of the esophagus or from alterations of the stomach itself

The human stomach is located beneath the diaphragm, through which the lower, terminal end of the esophagus passes The stomach appears as a dilated tube continuous with the distal end of the esophagus The upper curvature of the stomach is usually above and to the left of the *esophageal* orifice This expanded upper portion is the fundus and is commonly filled with air or gas The body (corpus) of the stomach is directed toward the attenuated right extremity or pyloric region and is subject to variations in size and shape depending upon functional activities, habits, disease, and volume of the contents The pyloric walls are marked by the heavy sphincter muscle which controls the

the lini

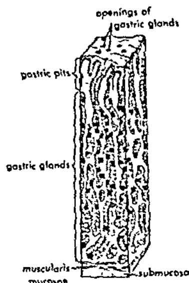
lined wi

In th ...rdiac and pyloric areas the glands

... of each gland The lower portion of the gland tubules is *lined* with chief cells which secrete zymogenic precursors, notably pepsinogen Scattered among the chief cells are the parietal, or border, cells which elaborate the hydrochloric acid and water of the stomach.

These glands may be activated directly by *the* presence of food, through stimulation by various nerves, or by stimulation from hormones released under both physiological and pathological conditions.

Three other layers are present in the stomach in addition to the *mucosa* The submucosa consists of a thin layer of loose connective tissue which lies



Semidiagrammatic view of a portion of *the gastric* mucosa showing gastric glands (From A. A. Maximow and W. Bloom, A Textbook of Histology, 6th ed., Saunders, 1952)

between the *mucosa* and the muscular layer. The latter is composed of three sets of muscles, an outer *longitudinal*, a middle circular, and an inner oblique. Surrounding the stomach and holding it and its vessels and nerves in place is the external serous layer which represents reflected peritoneum

The stomach temporarily holds food that is un-

...equent intervals until the stomach is emptied *See* DIGESTIVE SYSTEM. [E.C.ST.]

Stomach disorders

These include congenital defects, inflammations, tumors, and reactions to toxic materials and foreign bodies, as well as *minor* nonspecific complaints such as indigestion.

Congenital defects are infrequent. The most

com

bul

cor

the ... of the intestine, the duodenum. Occasionally, the stomach and

inte

mat

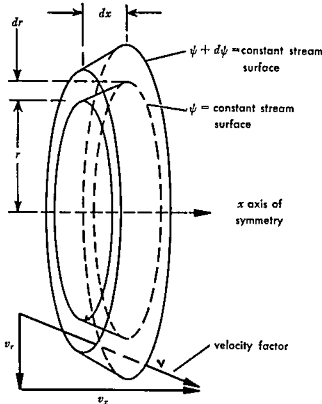
def

fror ...akness or excessive strain.

Inflammations vary from transient indigestion to full-blown hemorrhagic gastritis

... poisoning ...ges in the mucosal lining range from slight congestion and edema to extensive ulceration with massive hemorrhage.

Chronic gastritis is found in association with a shrunken (atrophic) mucosa and is most often



Section of axially symmetric fluid-flow field

Similarly on an r = constant cylinder, the mass flow between two stream-surfaces with axial distance dx between them is

$$2\pi\, d\psi = 2\pi \frac{\partial \psi}{\partial x} dx = -(2\pi r\, dx)\rho v_r$$

Thus the radial and axial velocity components at any point are

$$v_r = -\frac{1}{\rho}\frac{1}{r}\frac{\partial \psi}{\partial x} \qquad v_x = \frac{1}{\rho}\frac{1}{r}\frac{\partial \psi}{\partial r}$$

where ρ is fluid density. If the fluid motion is slow enough to neglect compressibility, ρ = constant and may be dropped from the above relations; ψ then measures the volume rate of flow inside a stream surface. *See* FLUID-FLOW PRINCIPLES

[A E BR.]

Stokes' theorem

the boundary curve of S, provided the sense of transcription of C is right-handed relative to $\mathbf{v}$ In symbols

$$\iint (\nabla \times \mathfrak{F}) \cdot \mathbf{v}\, dS = \int \mathfrak{F} \cdot \boldsymbol{\tau}\, ds$$

Here $\mathfrak{F}$ is a vector function, $\mathbf{v}$ is one of the two unit normals to the two-sided surface S, s is arc length measured positively in the sense which is right-handed relative to $\mathbf{v}$, and $\boldsymbol{\tau}$ is the unit tangent vector to C in the sense of increasing s. If $\mathbf{r}$ is the radius vector $\overrightarrow{OP}$ and $\mathbf{r} = \mathbf{r}(s)$ is the vector equation of C, then $\boldsymbol{\tau} = d\mathbf{r}/ds$ Further, if $\mathbf{r} = \mathbf{r}(u,v)$ is the equation of S, $\mathbf{r}_1 = \partial\mathbf{r}/\partial u$, $\mathbf{r}_2 = \partial\mathbf{r}/\partial v$, and $\mathbf{r}_1 \times \mathbf{r}_2 \neq 0$ (by assumption), then the vector $\mathbf{v}$ defined by $\mathbf{r}_1 \times \mathbf{r}_2 / |\mathbf{r}_1 \times \mathbf{r}_2|$ is a unit normal. The transcription Δu^+, Δv^+ Δu^-, Δv^- of a u,v-mesh is right-handed relative to $\mathbf{r}_1 \times \mathbf{r}_2$

By way of illustration, let S be the points of the plane

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j}$$

where $u = x$, $v = y$, and the points S are either on the circle

$$\mathbf{r} = a[\cos(s/a)\mathbf{i} + \sin(s/a)\mathbf{j}]$$

or in its interior. Let

$$\mathbf{v} = \mathbf{r}_1 \times \mathbf{r}_2 = \mathbf{i} \times \mathbf{j} = \mathbf{k} \qquad \mathfrak{F} = \tfrac{1}{2}(-y\mathbf{i} + x\mathbf{j})$$

then $\nabla \times \mathfrak{F} = \mathbf{k} \qquad \nabla \times \mathfrak{F} \cdot \mathbf{v} = \mathbf{k} \cdot \mathbf{k} = 1$

and $\iint \nabla \times \mathfrak{F} \cdot \mathbf{v}\, dS = \pi a^2$

Also $d\mathbf{r}/ds = \boldsymbol{\tau} = -\sin(s/a)\mathbf{i} + \cos(s/a)\mathbf{j}$

and on C $\mathfrak{F} = -\tfrac{1}{2}a \sin(s/a)\mathbf{i} + \tfrac{1}{2}a \cos(s/a)\mathbf{j}$

and $\mathfrak{F} \cdot \boldsymbol{\tau} = \tfrac{1}{2}a$

Consequently

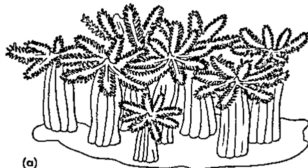
$$\int \mathfrak{F} \cdot \boldsymbol{\tau}\, ds = \tfrac{1}{2}a 2\pi a = \pi a^2 = \iint \nabla \times \mathfrak{F} \cdot \mathbf{v}\, dS$$

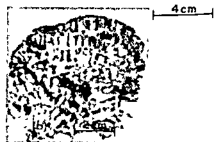
See CALCULUS OF VECTORS [H.V C.]

Stolonifera

An order of the Alcyonaria which lacks coenenchyme. They form either simple (*Cornularia*) or rather complex colonies (*Tubipora*). The polyp has a cylindrical body with a retractile oral portion which can withdraw into a solid anthostele or calyx protected by many calcareous spicules *Cornularia*, with a horny investment, lacks spicules. The base of the mature polyp is attached to a creeping stolon which is a ribbonlike network or thin flat mat from which daughter polyps arise. Daughter polyps never bud from the wall of the primary polyp Each polyp is connected by solenial tubes of the stolons, or by transverse platforms as in *Tubipora*.

The Stolonifera includes *Clavularia*, *Cornularia*, and *Tubipora*. The last is the organ-pipe coral with a dull red color, often discovered on coral reefs The stolons are transformed into skeletal tubes by





Stolonifera. (*a*) *Clavularia garciae* (after Y. Delage). (*b*) Skeleton of *Tubipora musica*

usion of the spicules These tubes are united by transverse stolons or solid platforms at several levels. See ALCYONARIA. [K.A.]

Stomach

The tubular or saccular abdominal organ of the digestive system adapted for temporary food storage and preliminary stages of food breakdown.

In lower vertebrates, such as fishes and amphibians, the stomach is little more than a simple tube quite similar to other portions of the gastrointestinal tract Small pockets, or diverticula, first appear in the reptiles See GASTROINTESTINAL TRACT.

In birds the stomach consists of a proventriculus and a gizzard The former is well supplied with glands which secrete softening and digestive materials, the latter is a strong muscular grinding organ whose action is often enhanced by the ingestion of small stones.

Mammals have stomachs which vary considerably in structure, according to the class or order. Although a single chamber is most common, some mammals, such as the ungulates, have as many as four. These may b...

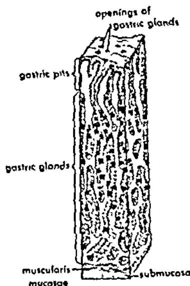
... the stomach appears as a dilated tube continuous with the distal end of the esophagus The upper curvature of the stomach is usually above and to the left of the esophageal orifice. This expanded upper portion is the fundus and is commonly filled with air or gas The body (corpus) of the stomach is directed toward the attenuated right extremity or pyloric region and is subject to variations in size and shape depending upon function...
ume o...
by th...
passa... of chyme into the duodenum

The stomach is lined with a mucous membrane that is usually thrown into longitudinal folds called rugae Most of the surface is covered with mucus-secreting epithelial cells but scattered throughout the lining are many small glandular pits which are lined with one or more types of secreting cells

In the cardiac and pyloric areas, the glands are primarily mucus-secreting In the fundus and body of the stomach however, the mucous cells occupy only the narrow neck of each gland The lower portion of the gland tubules is lined with chief cells which secrete zymogenic precursors, notably pepsinogen Scattered among the chief cells are the parietal, or border, cells which elaborate the hydrochloric acid and water of the stomach.

These glands may be activated directly by the presence of food, through stimulation by various nerves, or by stimulation from hormones released under both physiological and pathological conditions

Three other layers are present in the stomach in addition to the mucosa The submucosa consists of a thin layer of loose connective tissue which lies



Semidiagrammatic view of a portion of the gastric mucosa showing gastric glands. (From A. A. Maximow and W. Bloom, *A Textbook of Histology*, 6th ed., Saunders, 1952)

between the mucosa and the muscular layer. The latter is composed of three sets of muscles, an outer longitudinal, a middle circular, and an inner oblique Surrounding the stomach and holding it and its vessels and nerves in place is the external serous layer which represents reflected peritoneum.

The stomach temporarily holds food that is undergoing the mechanical and chemical changes of preliminary digestion A semiliquid fluid called chyme is the product of this action and is passed into the duodenum in small spurts which occur at frequent intervals until the stomach is emptied See DIGESTIVE SYSTEM [E.G.ST]

Stomach disorders

These include congenital defects, inflammations, tumors, and reactions to toxic materials and foreign bodies, as well as minor nonspecific complaints such as indigestion.

Congenital defects are infrequent The most common is pyloric steno...
bulblik...
constri...
the fir... part of the intestine, the duodenum. Occasionally, the stomach and other organs herniate into the ab...
matu...
defec...
from ... weakness or excessive strain

Inflammations vary from transient indigestion to full-blown hemorrhagic gastritis Acute gastritis may follow dietary indiscretions or the ingestion of alcohol, spices, acids, and many toxic substances, including those encountered in food poisoning. Changes in the mucosal lining range from slight congestion and edema to extensive ulceration with massive hemorrhage.

Chronic gastritis is found in association with a shrunken (atrophic) mucosa and is most often

seen in pernicious anemia and in aged patients. Low acid production is found on gastric analysis. Another poorly understood form of chronic inflammation is hypertrophic gastritis in which the mucosa becomes thickened, granular, and velvety, and shows accentuated folds. Hyperacidity is present and may well account for most symptoms.

Gastric ulcers are the basis for a large percentage of stomach complaints and may be acute or chronic. *See* PEPTIC ULCER.

Over two-thirds of all stomach tumors are carcinomas and account for one of the highest incidences of human malignancy, especially in men since they are affected twice as often as women. Carcinoma occurs most frequently in persons over 40 years of age. Atrophic gastritis and polyps are thought to be related, or at least appear to be predisposing diseases. One-half of stomach cancers are found in the lower stomach (the prepyloric and pyloric regions) and three general types are recognized, based on gross appearances, the ulcerative, the polypoid (or fungating), and the diffuse. Gastric carcinoma may lead to many complications, notably obstruction, hemorrhage, nutritional derangement, and prolonged vomiting with resulting dehydration and alkalosis. Perforation into the abdominal cavity is not uncommon and extension of the malignancy to adjacent and distant organs is the rule if diagnosis and surgical intervention are delayed. *See* ONCOLOGY.

Other malignancies of the stomach include leukemic and lymphoma infiltrations and sarcoma Spread of cancer from other primary sites to the stomach is seldom seen.

Benign tumors are uncommon, but myomas, neuromas, and polyps have the highest incidence.

Foreign bodies in the stomach are legion in numbers and variety. In most cases, small objects pass through to the bowel; large objects are retained, usually as chronically irritating masses; sharp objects may lacerate or even perforate the stomach wall. Bezoars are conglomerations of foreign material, often hair, which accumulate in the stomach and sometimes reach immense size

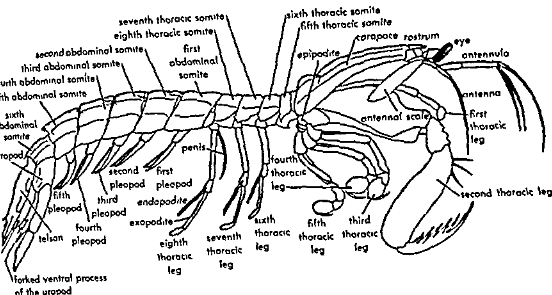
Gastric hemorrhage and gastric obstruction are the two most common and serious complications of many stomach disorders. *See* STOMACH. [E.C.ST.]

Stomatopoda

The only order of the superorder Hoplocarida belonging to the subclass Malacostraca of the class Crustacea. This order of the mantis shrimps contains a single family, the Squillidae with 8 genera (*Squilla, Pseudosquilla, Hemisquilla, Lysiosquilla, Coronida, Coronidopsis, Odontodactylus,* and *Gonodactylus*), and almost 200 known species. Both the Decapoda and the Stomatopoda are among the larger Crustacea. The size of adult Stomatopoda ranges from about 20 mm, in some *Gonodactylus* species to over 300 mm, in some species of *Squilla* and *Lysiosquilla.*

Morphology. The bodies of stomatopods are narrow and elongate, almost eruciform Only part of the cephalic and thoracic somites are fused and covered by the dorsal shield, the carapace. Those cephalic somites which bear the antennulae and the eyes are free and visible anterior to the carapace, and the last four thoracic somites are similarly exposed. Consequently there are 10 visible somites, 4 thoracic and 6 abdominal, between the carapace and the tail fan. The latter consists of a well-developed, sometimes peculiarly sculptured or deformed median plate, the telson, and the two uropods. Anteriorly, the carapace bears a flattened movable plate, the rostrum. The eyes are large, stalked, and movable. The antennulae have a 3-segmented stalk and three flagella, the antennae a 2-segmented stalk which bears a single flagellum and a large oval scale. The mouthparts consist of (1) a pair of mandibles which are strongly calcified and sometimes carry a palp, (2) a pair of maxillulae, small flattened organs formed by two segments and a small palp, and (3) a pair of maxillae, which are also flattened plates, but much larger and consisting of four segments. Of eight pairs of thoracic appendages the first is narrow, slender, and hairy, probably being used for cleaning purposes. The second thoracic leg is very strong and heavy. It has become a large raptorial claw, the distal segment of which folds back against the penultimate as does the blade of a pen knife. This claw shows a great resemblance to that of the praying mantis and for this reason the Stomatopoda are given the name mantis shrimps. The Stomatopoda catch their prey with their raptorial claws. In several species the last segment of the claw bears teeth on the inner margin, and the opposite margin of the penultimate segment often possesses fixed and movable spines, which evidently serve to make a better hold on the prey. The third to fifth thoracic legs are shorter and more slender than the second They end in short, more or less oval chelae which are similar to that of the raptorial claw. These legs are used for cutting up the food and for bringing it to the mouth. No exopodites occur on the first five pairs of thoracic legs, which may, however, carry a roundish flat epipodite at their base. The last three pairs of thoracic legs (sixth to eighth) are walking legs. They are elongate and have no chelae. They do not bear epipodites, but are biramous by the presence of exopodites. In the males a long slender penis is found on the inner side of the base of either of the legs of the last pair.

Each of the first five abdominal somites bears a a pair of pleopods which consist of a peduncular segment and two flattened blades, the exo- and endopod The outer blade or exopod bears a gill. The endopods of the first two pairs of pleopods of the male are transformed to a copulatory organ, the petasma. The appendages of the sixth abdominal somite, the uropods, consist of a peduncular segment on which are implanted a 2-segmented exopod and an endopod which consists of a single segment; furthermore, the peduncular segment bears a forked ventral process.



Schematic lateral view of a stomatopod

Distribution. The Stomatopoda are marine animals rarely found in brackish water. Most of the species are confined to tropical and subtropical areas, though some occur in the boreal and antiboreal regions The majority of stomatopods live in the littoral and sublittoral zones, but a few species have been found in greater depths, down to 760 meters.

Habitats. Most of the adult mantis shrimps are burrowing forms Many species of *Squilla* and *Lysiosquilla* dig their burrows in a bottom of sand or sandy mud Species of *Gonodactylus* have been observed to live in holes in coral reefs

Reproduction. After oviposition, the females carry the egg mass between the anterior thoracic legs until the larvae hatch The larvae differ considerably from the adults and are pelagic. They pass through a number of pelagic stages before starting their bottom dwelling life The difference between the larvae and the adults is so strong that the larvae formerly have been described as distinct species and genera such as *Alima*, *Erichthus* and others At present the identity of these "larval genera" with the [illegible] with the corresponding adult forms

Physiology. Species of *Squilla* produce a stridulating sound by rubbing the uropods over the lower surface of the telson *Gonodactylus* can make a snapping sound with the raptorial claws

The Stomatopoda are carnivorous and feed mainly on other crustaceans and fish, which they catch by a very fast movement of their raptorial claws Mollusks and other invertebrates are also eaten The stomatopod larvae also feed on living prey, largely small crustaceans.

Economic importance. The Stomatopoda are insignificant from an economic point of view. In some regions the larger species of *Squilla* and *Lysiosquilla* are eaten, but as a rule they are not highly esteemed

Fossils. The oldest fossil form which is recognized with certainty as a stomatopod dates from the Jurassic. Fossil stomatopods are rare. *See* MALACOSTRACA. [L.B.H.]

Stone and stone products

The term stone is applied to small fragments of rock and particularly to blocks or pieces of rock which are broken for use Certain geological and physical properties are essential if rock is to be used for commercial stone Generally a descriptive term, naming its specific use or property, accompanies the word stone as for example, building, roofing, crushed, or precious stone *See* GEM, ROCK

Dimension stone. The term dimension stone is applied to blocks of specified sizes, shapes, and surface finishes in contrast to the irregular fragmentary form of crushed stone Dimension stones are chiefly rectangular but may be in the form of columns or spindles Some dimension stones are carved into statuary or various structural and ornamental forms

Varieties used Granites, limestones, sandstones, and marbles are widely used as dimension stone. Basalts, diabases, and other dark igneous rocks are employed less extensively Soapstone is used to some extent

Requisite qualities Only a mere fraction of the rock occurrences on the face of the earth have the exacting requisite qualities demanded for dimension-stone uses. Stones may be sized, shaped, carved, or polished, but their inherent properties cannot be changed Obtaining stones of proper color, texture, or composition is purely a matter of selection The stone must be obtainable in large, sound blocks free from incipient cracks, seams, disfiguring blemishes, and mineral grains that might cause stains as a result of weathering. It

must have an attractive color, and generally a uniform texture, although clouded or banded stones may be preferred at times. Favorable acceptance by builders and architects is a requisite for successful development of a dimension-stone enterprise.

Uses. The principal uses of dimension stone are for building and ornamental applications. Years ago, solid stone walls of buildings were common, but since the advent of reinforced concrete for bearing walls, stone has been used as a veneer only. It may be used for entire exteriors or for base courses, sills, caps, or trim of buildings made of brick, wood, or other materials. It furnishes a permanent finish of architectural dignity and minimum upkeep The more ornamental types of marbles, onyx marbles, travertine, colored limestones and sandstones, and polished granites are employed extensively in the interiors of public buildings for paneling, wainscoting, floor tile, stair treads, columns, spindles, and other structural and decorative units.

Another important use is for memorials, ranging from simple markers and headstones to elaborate monuments and mausoleums. Marbles and granites that normally take a good polish are the stones used most extensively for memorials. Dimension stone is prepared for its many structural and ornamental uses in mills equipped with gang saws, diamond saws, planers, rubbing beds, turning lathes, polishing machines, and other equipment similar in some respects to that found in metal- and wood-working shops

Stone used for flagging consists of thin slabs that may be rectangular or irregular in shape. The stones most commonly used are thin-bedded sandstones and quartzites that have closely spaced natural partings and slates that split readily into thin slabs. Flagging is used for stepping stones, walkways, and terraces, and for flooring patios and porches Attractive patterns may be made by combining stones of different colors

The manufacture of curbstones for use along streets and highways was an important use for granites, quartzites, and the harder sandstones many years ago, but concrete has replaced them extensively. Because of their high resistance to the chemical action of de-icing compounds, the natural stones are preferred in many northern localities.

Sandstones were formerly widely used for abrasive purposes Pulpstones, grindstones, whetstones, and similar products made of sandstone have, however, been replaced extensively by synthetic abrasives. An abrasive use for quartzite that persists is for making liners and balls for tube and ball mills employed for grinding mineral products into fine powders. Because of their high melting point (about 3000°F) sandstone blocks known as firestones are used for metallurgical furnace linings.

Granites and diabases are used for making special plates with true, uniform surfaces (surface plates) for the precise testing of instruments.

Soapstone is used for building purposes, for acidproof laboratory equipment, for laundry tubs and aquariums, and for chemical-tank linings. The harder varieties are employed for floor tile and stair treads. *See* TALC.

Slate (roofing stone). Commercial slate must be uniform in quality and texture, reasonably free from knots, streaks, or other imperfections, and have good splitting properties. *See* SLATE

Roofing slates are important products of most slate quarries. However, the roofing-slate industry has declined considerably because of competition from other types of roofing. Slate is also used for milled products such as blackboards, electrical panels, window and door sills and caps, base boards, stair treads, and floor tile Laboratory table tops, sinks, and surface plates are also made of slate. Slates can be split into sheets of any desired thickness whereas other varieties of stone except flagging are cut into thin slabs by sawing

Crushed and broken stone. All the principal types of stone—granite, diabase, basalt, limestone, dolomite, sandstone, and marble (except schist and slate)—may be used as sources of commercial crushed stone. Limestone is by far the most important; it constitutes about 75% of all the crushed stone sold.

Requisite qualities. Sound, hard stone, free from surface alteration by weathering, is demanded Stone which breaks in chunky, more or less cubical fragments is preferred. Flat or splintery fragments like those obtained from schists or slates are undesirable, and such rocks are used sparingly if at all for crusher products. Commercial stone should be free from certain deleterious impurities, such as opalescent quartz, which may react with lime in cement and cause disintegration of concrete in which the stone may be used as aggregate. Crushed stone should be free from associated clay or silt It is sometimes cleaned by washing.

Uses Large blocks or irregular masses of stone known as riprap are used widely for shore protection along rivers and harbors or to make spillways at dams. Waste blocks from dimension-stone quarries are commonly so used. Stone also is crushed and pulverized for many uses The amount of crushed stone used annually now exceeds that used for dimension stone.

Crushed stone. Many millions of tons of stone are crushed and screened, chiefly in sizes ranging from ½ to 3 in and are used in various branches of the construction industries The preparation of crushed stone involves blasting rock ledges with high explosives and loading the fragments into cars or trucks for haulage to massive crushers that reduce the blocks to smaller sizes Smaller crushers are used for further reduction, and the fragments are classified according to size by screening *See* QUARRYING.

The principal uses are as concrete aggregate, as road stone, or as railway ballast. Another important use for limestone is as a fluxing material to remove impurities from ores smelted in metallurgical furnaces. Dolomite, either [illegible] aw state or

lcined at a high temperature (dead-burned), is ed as a refractory for furnace linings Limestone also used very extensively in such chemical and rocessing industries as alkali, calcium carbide, lass, paper, paint, and sugar manufacture. It is lso used for filter beds and for making mineral ool. Stone chips are used for terrazzo flooring and tucco, and granular material for poultry grit. *See* CONCRETE; REFRACTORY

Pulverized stone. Large quantities of stone, prin-cipally limestone, are pulverized and sold as fine

tilizers, paints, road asphalt mixtures, and roofing mastic. White stone powders are used for coal-mine dusting to prevent explosions Limestone reduced to an impalpable dust is used as a whiting substitute in putty, calcimine, floor coverings, tooth paste, pottery, plastics, explosives, and numerous other products.

Raw material for special uses. Limestone is the principal raw material used in making portland cement, a product that has attained an annual value in the United States exceeding $900,000 000 Cement has a multitude of uses, particularly in the building trades and in highway and engineering construction *See* CEMENT

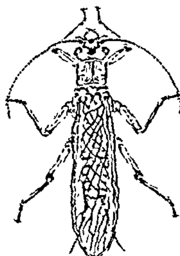
Limestone is the sole raw material used in making lime. Lime manufacture is a calcining (heat-treating) process during which the carbon dioxide is driven from the calcium carbonate (limestone) leaving the calcium oxide (lime). Lime is used extensively in agriculture for soil amendment and in the building trades for making plaster and mortar (building lime) but its principal use is in a great variety of chemical and industrial processes such as water purification, sewage treatment, road stabilization, salt and petroleum refining, and in the manufacture of paper, sugar, glass, rubber, food products glue, soap, and many other products *See* LIME (INDUSTRIAL)

For a discussion of the compressive (crushing) strength of rocks *see* ENGINEERING GEOLOGY

[O B]

Stone fly

Any member of the insect order Plecoptera The Plecoptera are characterized by two pairs of membranous, net-veined wings, the hind pair much larger than the front pair, folded in plaits and lying flat on the back when at rest, two pairs of long abdominal cerci, long antennae; incomplete metamorphosis; and chewing mouthparts usually vestigial but sometimes well developed These drab, feeble fliers are seldom found far from water The nymphs resemble the young of damsel flies, but have two cerci rather than three tracheal gills at the end of the abdomen Most stone flies do not eat as adults, mating and dying soon after emergence. The nymphs develop in 1–2 years, usually in swift, well-aerated water There are about 1500 species, 300 of which occur in North America They are of some value as fish food *See* PLECOPTERA. [J D B]



The stone fly, *Perla* sp. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Storage battery

An assembly of identical voltaic cells in which the electrochemical action is reversible so that the battery may be recharged by passing a current through the cells in the opposite direction to that of discharge While many nonstorage batteries have a reversible process, only those that are economically rechargeable are classified as storage batteries. *See* BATTERY (ELECTRIC); PRIMARY BATTERY.

Storage batteries, sometimes known as electric accumulators or secondary batteries, have two general classifications, lead acid and nickel-alkaline Active materials and electrolytes for both classes of batteries will be explained in subsequent paragraphs

Storage batteries are used for diesel engine and gasoline-engine starting, for standby in areas served by dc power, for circuit breaker closing and tripping, and in marine power systems, railway-car lighting telegraph and telephone services, emergency lighting and power services, photographic and sound services, farm lighting, electric-vehicle propulsion, and in many military applications.

LEAD–ACID STORAGE BATTERY

The lead-acid type storage battery is so classified because the electrolyte is an acid and the plates are largely lead. Active materials for the positive electrodes, or plates, are lead, lead antimony, or lead calcium, depending on the service or application Negative plates usually have a lead paste with an inert material added and are thus somewhat different in composition from the positive plates

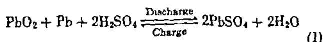
The electrolyte is sulfuric acid diluted with water Specific gravity (sp gr) at full charge is between 1.200 and 1 300 for new cells or batteries: 1 200–1.210 for high temperature locations, 1 200–1.220 for stationary batteries, 1 280 for highway and industrial trucks, locomotives, and automo-

biles, and 1.300 for special applications. Ordinarily 1.180 is an indication that the battery requires recharging. Specific gravity is measured by a hydrometer. *See* HYDROMETER.

The lead-acid battery maintains a preeminent place among all commercial types of storage batteries in volume of manufacture.

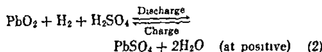
Principles of operation. A great many types of lead-acid cells are produced, but all have certain features in common. One is the open-circuit cell electromotive force (emf), which exists between a positive lead peroxide (PbO_2) electrode and a negative sponge lead (Pb) electrode when the two are dipped into sulfuric acid electrolyte ($H_2SO_4 + H_2O$). *This value is independent* of the quantities of lead peroxide, lead, or electrolyte but does vary with temperature and sulfuric acid (H_2SO_4) concentration. At 25°C the emf varies from 2.050 volts with acid at 1.200 sp gr to 2.148 volts with acid at 1.300 sp gr. The relatively small variation with temperature is given in millivolts per °C over a range 0–40°C, as 0.30 for 1.200-sp gr electrolyte, 0.22 for 1.250 sp gr, 0.19 for 1.280 sp gr, and 0.18 for 1.300 sp gr.

The reversible equation below represents the cell reactions in so far as beginning and end materials are concerned. It is known as the double-sulfate theory, since lead sulfate ($PbSO_4$) is formed at *both electrodes.*

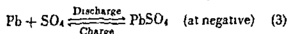


$$PbO_2 + Pb + 2H_2SO_4 \underset{\text{Charge}}{\overset{\text{Discharge}}{\rightleftharpoons}} 2PbSO_4 + 2H_2O \qquad (1)$$

The above can be split into two equations indicating the reactions at the two electrodes:



$$PbO_2 + H_2 + H_2SO_4 \underset{\text{Charge}}{\overset{\text{Discharge}}{\rightleftharpoons}} PbSO_4 + 2H_2O \quad \text{(at positive)} \qquad (2)$$



$$Pb + SO_4 \underset{\text{Charge}}{\overset{\text{Discharge}}{\rightleftharpoons}} PbSO_4 \quad \text{(at negative)} \qquad (3)$$

The above equations satisfy the amounts of reactants involved but do not indicate the ionization and intermediate steps. On discharge the over-all effect is a reduction of PbO_2 at the positive electrode and an oxidation of Pb at the negative electrode, accompanied by sulfation in both cases. In charging, a counter voltage is imposed on the cell terminals, and current is forced through the cell in a direction opposite to that in which the cell discharges. This reverses the ionic movements in relation to the electrodes and, in effect, reverses the cell reactions. On discharge the electrolyte specific gravity decreases, and on charge it increases. Specific gravity serves as an index of state of charge.

Reactants. For a given quantity of electricity, such as ampere-hours, the three reactants PbO_2, Pb, and H_2SO_4 take part *in the reaction in* amounts governed by Faraday's law. Thus, for a 1 ampere-hour discharge, 3.866 grams (g) of sponge lead (Pb) is converted to $PbSO_4$, 4.463 g PbO_2 is converted to $PbSO_4$, and 3.660 g of H_2SO_4 is consumed.

A cell constructed to contain exactly the amounts of reactants given above, however, would not yield 1 ampere-hour of capacity even under optimum conditions. Action at each electrode is slowed drastically when the concentration of H_2SO_4 in the electrolyte approaches a low figure. This is particularly true of the positive plate where, as can be observed from Eqs. (2) and (3), there must be excess acid present around the electrode to consummate the reaction. The final consumption of SO_4 by the positive plate, however, is no more than that consumed by the negative electrode. But even if ample H_2SO_4 were present, 1 ampere-hour of capacity still would not be attained, since there will always remain an appreciable amount of PbO_2 or Pb, or both, which cannot be reached by the electrolyte. The capacity attained in relation to what should be obtained theoretically from the amount of reactants present is known as the utilization coefficient. This coefficient varies with types of cells, rate of discharge, and temperature. Unfortu-

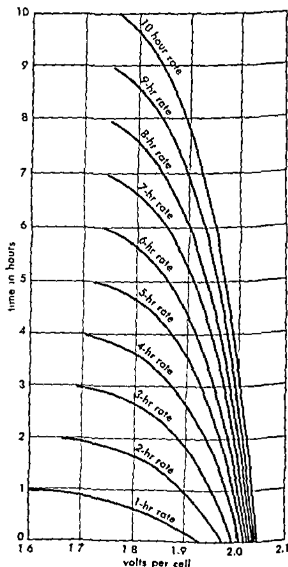


Fig. 1. Typical volt-time characteristics for a lead-acid cell for various discharge rates. (*Electric Storage Battery Co.*)

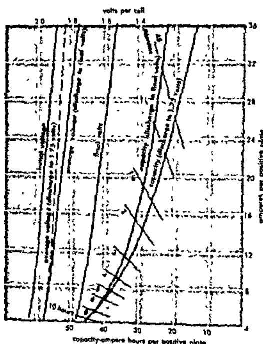


Fig 2 Rated-discharge characteristics for a lead-acid cell (Electric Storage Battery Co.)

nately, it is a low value even under best conditions.

In addition, there are circumstances which prevent full utilization of the active materials as the discharge progresses One of these is the coating of nonconducting sulfate that forms on the active materials. Another is the diminishing conductivity of the electrolyte as the H_2SO_4 content decreases. A third is the decrease in porosity as sulfate forms in the pores of the materials, in turn impeding diffusion of H_2SO_4 This latter effect is particularly serious at high-current rates, since at higher current densities the tendency to concentrate reaction at the surface of the plates is greater As a result, the pore openings at the plate surfaces become blocked with sulfate, restricting conduction and diffusion to the interiors of plates Plates destined for high discharge current densities are therefore made relatively thin. By substituting many thin plates for a few thick plates containing the same amounts of active materials, the utilization coefficient at high rates, and hence the capacity attainable, will be increased.

Figure 1 shows a typical set of volt-time curves for a lead acid cell, illustrating the variations from a 1-hour rate at high discharge current to a 10-hour rate at low discharge current. Figure 2 illustrates the decrease in ampere-hour capacity with increase in discharge current.

If a short time for diffusion is allowed after a high rate discharge, more of the unused possible capacity of the cell will be made available

Cell temperature has an appreciable effect on capacity, largely because the viscosity of the electrolyte changes Thus the diffusion of H_2SO_4 is retarded at low temperatures.

As already pointed out, there is a minimum acid concentration permissible for completion of reactions Also, experience has shown that negative plates do not function well if the full-charge specific gravity is over 1.300, although positive plates operate more efficiently in high specific gravity. The usual range of full-charge specific gravity is 1.200–1.280, the choice depending on the applica-

[illegible]

eration must be given to the kind of service for which a cell is destined, second consideration to design features that eliminate operational troubles. A compromise, such as between life and weight or life and cost, is usually required.

Pasted plates. In the most familiar type of plate construction, a grid, such as the ladder type illustrated in Fig. 3, is pasted with a lead oxide, acid, and water mix. This is followed by a processing which finally converts the active material to PbO_2 for positive plates (chocolate-brown color) and to sponge lead for negative plates (grey color).

Negative plates are always of this type when used with pasted positive plates or positive plates

[illegible]

thickness will probably be intermediate; if for stationary service, such as standby or with only occasional discharge, the plates will be thick; if for signal or alarm service and where charging is done only at long intervals, they will be very thick and have lead or lead-calcium grids.

Manchester plate. These plates consist of a heavy alloy grid with circular openings into which pure lead "buttons" are pressed. These buttons are made from lead tape by crimping and rolling to develop a large surface area (see Fig. 4). A forming agent in dilute sulfuric acid electrochemically forms a layer of PbO_2 on the surface of the button. Manchester plates are usually mounted in a cell with pasted negatives and in a relatively large quantity of low-gravity acid.

The cells are heavy and bulky. They are used in stationary installations, as for telephones, switch

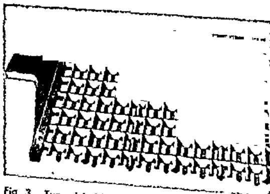


Fig 3. Typical ladder-type grid showing a portion of it pasted. (Electric Storage Battery Co)

operation, and large emergency lighting, where they are "floated" on a line of constant voltage or trickle-charged with a constant current and are only occasionally discharged. In such service Manchester plates give exceptionally long life.

Gould spun plates. This type of positive plate, shown in Fig. 5, is manufactured from heavy sheet lead by passing the plate between disks which cause the lead to flow in between them to form leaves and spaces. After PbO_2 is formed on this developed surface, the plates are assembled with pasted negative plates and used in substantially the same types of service as Manchester plates. An advantage of this plate is elimination of antimony, hence local action, from the cell construction. This advantage is usually gained at some sacrifice of life.

Exide Ironclad plate. In this positive plate the active material is held in a porous-walled tube with a central alloy spine as conductor. The tube is made of glass fiber armored with perforated thin plastic.

This plate is used in many applications but is particularly successful where the service calls for repeated or routine deep-discharge cycles, such as in industrial trucks and mine locomotives.

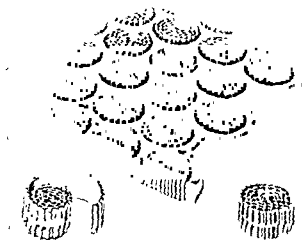


Fig. 4. Section of Manchester plate with detail of lead button. (*Electric Storage Battery Co.*)

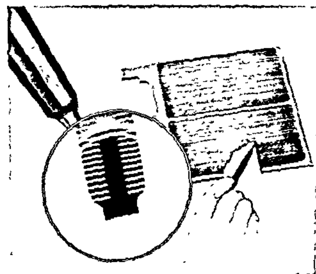


Fig. 5. Gould spun Planté positive plate. (*Gould-National Batteries, Inc.*)

Freezing of electrolyte. The freezing points of the usual range of sulfuric-acid electrolytes at full-charge specific gravities (from −52°C for 1.250 at 15°C to −70°C for 1.300 at 15°C) are well below most arctic temperatures, but the end-of-discharge specific gravities can result in freezing points above arctic temperatures unless precautions are taken. With the proper choice of separator, a high-gravity acid can be used in severe arctic conditions without detriment to the negative electrodes and yet can have a relatively high end-of-discharge gravity.

Charging. For fast, yet efficient and noninjurious, charging the modified constant-potential method is recommended. A high current rate is used until a voltage, such as 2.38, is obtained. This voltage is then maintained with a decreasing current until the finishing rate recommended by the manufacturer is reached. The finishing rate is continued to the end of the charge. Several methods are used for setting the end of the charge, the best known being arrival at a constant potential, arrival at a constant specific gravity, or charging a certain per cent of ampere-hours in excess of the ampere-hours that have been taken out. A lengthy but efficient charge can be made using the finishing rate from the start. A two-step charge can be made using first a high current and then the finishing rate, the change being made automatically by a voltage relay or ampere-hour meter.

Containers. Compositions of asphaltic or bituminous materials with inert fillers and fibers are used in some of the cheaper grades of batteries. Better grades use hard rubber, frequently blended with synthetic and reclaimed rubbers. These are used in airplanes, ships, and submarines as well as in automobiles. Glass and ceramic jars are confined almost entirely to stationary cells, where weight and volume are of minor importance and chance of breakage is small. Plastic containers are coming more into use, displacing many of the other types.

Battery troubles and remedies. Some of the more important battery troubles are corrosion of the grid, shedding of active materials, and self-discharge.

Corrosion. Gradual wearing away of the grid containing the positive active material results in subsequent disintegration of the plate. Cells subjected to repeated deep discharges and overcharges are particularly prone to this trouble. The antimony used in the grid alloy (5–12% Sb is used to impart stiffness and castability) aids in resisting corrosion under certain conditions. The addition of small percentages of arsenic and silver increases this resistance to a notable extent.

Shedding of active materials. This usually pertains to the positive active material of cells that are

subjected to overcharging, whereby gas formation strains and loosens particles of material near the surface of the plate. Unless retainers, such as slotted rubber or plastic or mats of glass fiber, are used against the positive plates, active material may drop to the bottom as sediment and cause short circuits between plates. The material may also be carried by gas streams to the top of the cell to pile up and short-circuit there.

Self-discharge or local action. Self-discharge of negative plates is caused by deposition of certain metals on the plate to form a voltaic couple with the sponge lead. Since these couples, actually shorted, have an emf in excess of the hydrogen overvoltage on the metal deposit, an evolution of hydrogen ensues, and the adjacent lead is sulfated. Metals most frequently producing local action include antimony, copper, silver, and, less frequently, tin, arsenic, bismuth, platinum, and nickel.

Other metals, such as iron and manganese, whose salts readily exist in solution in two stages of oxidation, can reduce the positive plate by diffusion or convection and oxidize the negative plate with $PbSO_4$ formation at both electrodes.

If self discharge of the negative electrodes becomes rapid or is allowed to act over a long period, or if a cell stands for some time in a discharged condition, the sulfate crystals become large, hard, and difficult to reduce to lead.

Undercharging. This causes buckling, or warping, of plates and sulfation. This condition is usually due to unequal work on the two sides of the positive plates, resulting from unequal electrolytic attack or unequal expansion of active materials.

Overcharging. This causes corrosion, buckling, washing and overheating. It is caused by high charging rates, which should be tapered off as the battery becomes charged.

Densification of negative material. A deficiency of certain organic and inorganic compounds, used as additives to the active material, permits coalescence of lead particles with a consequent loss of porosity. Compounds added to the negative plate material to prevent this are commonly called expanders.

Separator shorts. The separators between positive and negative plates may be oxidized through contact with PbO_2, resulting in puncture and short circuits between plates.

ALKALINE-TYPE STORAGE BATTERY

The alkaline-type storage battery is so classified because the electric energy is converted from chemical action through an alkaline solution. One type of battery has positive plates with an active material of nickel and negative plates of iron. Another type uses nickel and cadmium. A third uses silver oxide and zinc.

Nickel-iron alkaline cell. This battery is composed of cells having nickel and iron in an alkaline solution. It is also known as the Edison battery.

The positive active material in these cells is a higher oxide or hydroxide of nickel. The negative material is fine iron powder. The electrolyte is 1,200 sp gr at 60°F (15.56°C) potassium hydroxide KOH, to which a little lithium hydroxide is sometimes added. A freshly charged positive electrode contains some NiO_2, which is unstable and decomposes in a few days. As a result the freshly charged cell will have an emf of 1.48 volts, which decreases in a few days to a stable 1.35 volts. This unstable oxide (NiO_2) can also affect electrical performance, and the hours between full charge and start of discharge should be noted in stating capacities. The KOH electrol [illegible]

[illegible]

hence gives no indication of state of charge. The latter can be determined from the voltage, when discharged briefly through a certain fixed resistance and referred to a calibration. During discharge the positive material is reduced to nickelous hydroxide, and the iron in the negative is oxidized to ferrous hydrate. Figure 6 shows typical discharge curves of a 100-amp-hr nickel-iron cell.

The positive material is packed into perforated steel tubes with alternate layers of finely flaked nickel to provide conductivity. The tubes are then pressed and clamped into a steel frame. The negative material is packed into pockets of perforated steel shaped like long, thin boxes. These are also pressed and crimped into a steel frame. All steel parts, including the welded steel container, are nickel-plated. Positive and negative plates are shown in Fig. 7.

Manufacturers warn against operating nickel-iron cells above 115°F. Also, depending somewhat on discharge rate, capacities drop rapidly below a critical temperature of about 36°F.

In charging Edison cells a current not less than one-half the normal (5-hour) rate is used because a large quantity of hydrogen gas must evolve at the negative plate to effect the reduction to metallic iron.

Edison cells have found extensive application in industrial trucks, for car lighting, and for standby or emergency sources of power.

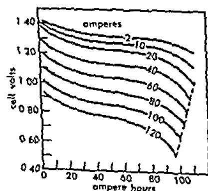


Fig. 6. Typical volt time curves of nickel-iron alkaline (Edison) cells for various discharge rates. (From A. E. Knowlton, ed., Standard Handbook for Electrical Engineers, 9th ed., McGraw-Hill, 1957)

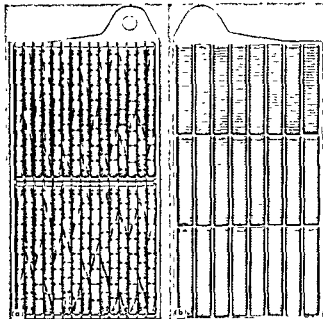


Fig. 7. Plate construction of nickel-iron (Edison) cell (a) Positive plate. (b) Negative plate. (*Thomas A. Edison Industries*)

Nickel-cadmium alkaline cells. In the original types of Ni-Cd (Jungner) cells the materials and structural features for the positive plates are quite similar to those described for the Ni-Fe cell. Differences are that graphite, instead of flake nickel, is used to give conductivity to the material and that some manufacturers use pockets instead of tubes. Negative plates of these original types also have the same structural design as those for the Ni-Fe cells, but the active material is cadmium instead of iron, and on discharge this goes to cadmium hydroxide.

Decomposition of some NiO_2 after full charge produces the same effects on open-circuit voltage and discharge immediately after charge as described for the Ni-Fe cell. Following this initial decomposition the emf will be close to 1.30 volts.

During World War II the Germans developed a sintered-plate type of Ni-Cd cell. Extremely fine nickel powder, obtained from decomposition of nickel carbonyl, is sintered in a mold around a nickel or nickel-plated screen. For positive plates these plaques are impregnated with a nickel salt (usually nitrate) and processed to throw out the nickel hydrate in the pores. Plaques for the negatives are impregnated with a cadmium salt (nitrate or chloride) and processed in a manner like that for the positive.

Sintered-plate cells are displacing the original types, being superior in several respects. They have less internal resistance and a higher utilization coefficient, and they perform better at both higher and lower temperatures.

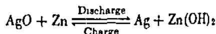
The electrolyte is a solution of potassium hydroxide (KOH) made with specific gravities rang- [illegible]
[illegible] several stationary
[illegible] and in aircraft.
[illegible]

Containers are made of nickel-plated steel or plastic.

Charging can be done rapidly and efficiently by constant-current, constant-potential, and modified constant-potential methods; gassing begins around 1.47 volts, and when using normal charge rate (5-hour), the end voltage will be 1.75.

Typical discharge and charge curves are shown in Figs. 8 and 9.

Silver oxide–zinc alkaline cell. Silver oxide positive plates and sponge-zinc negative plates came into use during the late 1940s. They have high ampere-hour and watt-hour capacities per unit of volume or weight. A high-specific-gravity KOH solution, up to 1.450, has been found advantageous in minimizing local action. Cell reactions can be expressed as



$$AgO + Zn \underset{\text{Charge}}{\overset{\text{Discharge}}{\rightleftharpoons}} Ag + Zn(OH)_2$$

Charging can be accomplished by a constant current or modified constant-potential charge, as long as the cell voltage does not exceed 2.1 volts at any time.

Typical discharge and charge curves are shown in Figs. 10 and 11.

Silver oxide–zinc cells are used both as primary and secondary cells for military use and for nonmilitary applications where battery power with minimum weight is an essential consideration.

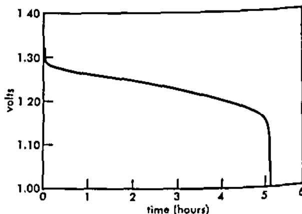


Fig. 8 Typical discharge curve for sintered-plate nickel-cadmium cell. (*Electric Storage Battery Co.*)

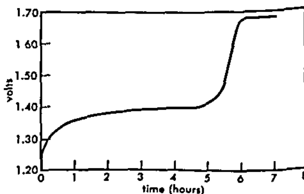


Fig 9. Typical charge curve for sintered-plate nickel cadmium cell. (*Electric Storage Battery Co.*)

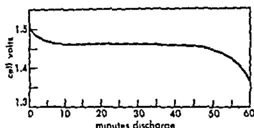


Fig 10 Typical discharge curve for silver oxide-zinc cell (Electric Storage Battery Co.)

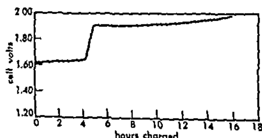


Fig 11 Typical charge curve for silver oxide-zinc cell (Electric Storage Battery Co.)

Freezing of alkaline electrolyte. The use of high-gravity KOH electrolyte for Ni Cd and AgO-Zn cells eliminates freezing under severe arctic conditions. High-specific-gravity electrolyte cannot be used with Ni-Fe cells.

Venting of storage cells. Venting must be provided for all storage cells to permit escape of local-action gas or gas generated in the charging process The only exceptions are the special sealed cells, in which gassing is held to a minimum and any hydrogen or oxygen generated is recombined through catalysis

Provision for the escape of gas has necessitated numerous devices to prevent spillage of electrolyte from cells in applications such as aircraft.

[H P.MU.]

Bibliography J. T Crennell and F. M. Lea, *Alkaline Accumulators*, 1928, A E Knowlton (ed), *Standard Handbook for Electrical Engineers*, 9th ed , 1957; G. W Vinal, *Storage Batteries*, 4th ed , 1955.

Storage devices

Devices capable of assuming distinguishable states, which when given a stimulus or input, assume another or the same state uniquely representative of that input and previous state, also called memory devices

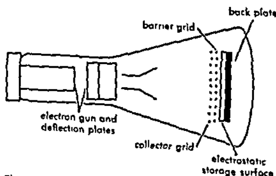
Storage devices are extensively used in switching circuits and in digital computers in the form of hold contact relays, electronic flip-flops, magnetic cores, and the like, either in control and gating applications or as arithmetical registers for storing digitally represented numbers

Other devices, used for bulk storage of numbers and other information, are punched cards, magnetic tape, punched-paper tape, and magnetic disks See DATA PROCESSING SYSTEMS.

For more rapid access to large amounts of information (1000 or more separate items or words), magnetic drums, electrostatic storage tubes, mercury delay lines, and magnetic-core matrices are used. For discussion of magnetic-core matrices, and also the use of storage devices in digital computers see DIGITAL COMPUTER.

Magnetic drum. This is a revolving cylinder on which binary digits 1 and 0 may be represented by magnetized spots. Tracks of spot positions around the circumference of the drum pass by magnetic heads, mounted on bars around the cylinder, for recording (writing) and sensing (reading). These spots may directly represent binary numbers or may represent decimal digits by appropriate coding A typical access time for reading a datum from the drum into a computer register is half a drum revolution, or about 5 milliseconds The drum is used either as the main memory of the machine or for auxiliary bulk storage.

Electrostatic storage. Binary digits can be stored in the form of small charged spots on the face of a cathode-ray tube. Spots are created by impingement of an electron beam on the storage surface, the beam being directed to various locations by horizontal and vertical deflection plates (*see* CATHODE-RAY TUBE). On a tube 3–4 in. in diameter, about 1000 binary digits can be recorded,



Electrostatic storage tube

and access to any bit is obtained in about 10 microseconds Electrostatic storage, as a truly fast-access storage device, has almost entirely given way to the magnetic core

Delay line. The delay-line storage device is commonly a column of mercury, along which mechanical waves travel at a frequency of 1 megacycle or more, the waves representing binary digits Electrical signals are introduced into a quartz crystal which, because of its piezoelectric property, sets up mechanical vibrations in the mercury column The waves are converted back into electrical signals by a second crystal at the other end of the column A delay line can store several hundred bits of information. *See* PIEZOELECTRICITY. [R J.N]

Bibliography: A. D Booth and K. H. V. Booth, ... R. K ... Cir.

Storage tube

An electron tube into which information can be introduced and then extracted at a later time; also called a memory tube.

Operation. The process of introducing information into a storage tube is known as writing, and that of extracting useful information is called reading. The deliberate removal of information from the storage surface is called erasing. In some tubes reading automatically effects erasing; in others a separate operation is required. A charge-storage tube is one in which information is retained on a surface in the form of electric charges.

The characteristics of a storage tube are largely governed by the nature of the storage surface This surface is usually a deposit or sheet of insulating or semiconducting material The point-by-point potential of this surface is varied, in a controlled way, by the dielectric charging processes associated with electron bombardment of an insulator or semiconductor. The polarity of the resulting potential pattern depends upon the beam-accelerating voltage employed and the voltages applied to the elements adjacent to the storage surface. The low conductivity of the storage-surface material ensures that the charge (or potential) pattern will not be dissipated before the reading operation is initiated and completed.

In the reading operation, the potential pattern established by writing controls either the percentage of the primary beam that is able to pass through the openings of a fine mesh screen on which the storage surface may be deposited or else the magnitude of the secondary-electron current, which results when the primary beam strikes the storage surface. In either case, the current collected by the output electrode is modulated in accordance with the stored charge pattern.

Storage tubes may be classified according to the general nature of the input and output signals associated with them. Thus there are visual-electrical, electrical-electrical, and electrical-visual storage tubes. Visual-electrical storage tubes are better known as camera tubes and are described elsewhere (*see* TELEVISION CAMERA TUBE) In the following section several commercially important cathode-ray charge-storage tubes in the latter two classes will be briefly discussed.

Ordinary cathode-ray tubes have been used as electrical-electrical storage tubes in the buffer stages of high-speed digital computers The phosphor screen and the envelope faceplate of these tubes serve together as the storage medium. An output electrode is provided by affixing a conducting membrane to the outside of the faceplate. Cathode-ray tubes especially designed for computer service are characterized by phosphor screens chosen for secondary-emission properties, stability, and purity, rather than for luminescent efficiency. These tubes are popularly called Williams tubes, after F. C. Williams, a pioneer in their application.

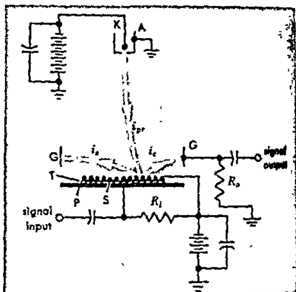


Fig. 1 Signal-converter storage tube (Radechon) (From M. Knoll and B. Kazan, Storage Tubes, Wiley, 1952)

Radechon. This vacuum tube, known also as a barrier-grid storage tube, is a single-gun device The dielectric storage medium in this case is usually, but not necessarily, a sheet of mica T sandwiched between a continuous metal backing plate P and a fine mesh screen S (see Fig. 1). The electron beam i_{pr} is formed by the electron gun represented by cathode K and anode A. The collector G collects reading current i_c. This current develops an output signal voltage across output resistor R_o. The screen, the barrier grid, serves to restrict undesirable redistribution of secondary electrons over the storage surface. Additionally, the potential applied to the barrier grid is that to which the storage surface charges under the influence of beam bombardment Writing consists of establishing localized voltage gradients, which may be of either polarity, across the storage layer. In the tube shown in Fig 1 these gradients are established by varying the backplate potential by means of a signal fed into the input resistor R_i The output current in reading is that which flows to the collector electrode when discharging this gradient

The Radechon is highly versatile, finding use in simple delay schemes, signal-to-noise improvement applications, signal-comparison service, and systems for the conversion of signal-time bases.

Recording storage tube. This tube employs a storage medium in the form of an insulating deposit T on one side of a fine mesh screen S (Fig. 2) In the reading mode, the scanning electron beam approaches the storage mesh screen at low velocity. The magnitude of the reading beam current which is able to penetrate the openings of the storage mesh screen and impinge on the output electrode is controlled by the potential pattern established on the insulating deposit. This pattern is related to the input signal which modulates the electron beam during operation of the tube in the

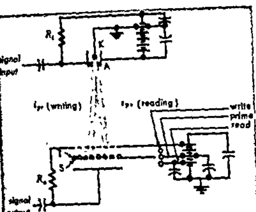


Fig. 2. Signal-converter storage tube (recording storage tube).

writing mode Both one- and two-gun versions of this device are produced, the latter for simultaneous writing and reading operation, which is required in some applications

This type of tube is particularly applicable to systems in which many faithful output copies of a stored pattern are needed, because the reading process is not an erasing process as well

Graphechon. This tube is fitted with two electron guns, one on each side of the storage medium, a thin film T of insulator or semiconductor deposited on a thin substratum of metal P which is supported by a fine mesh M (see Fig 3). The reading process, which is also an erasing process, attempts to maintain an equilibrium voltage gradient across the storage layer. This gradient is discharged point by point by the writing beam, which is sufficiently energetic to penetrate the storage layer as well as its supporting substratum This penetrating beam bombardment produces additional charge carriers which hasten the discharging process. The current that flows to the wire mesh, as the reading process strives to reestablish the equilibrium gradient, constitutes the output signal. The collector removes the return current from the storage medium.

The Graphechon is intended for scan-conversion service, typically converting radar PPI signals to standard TV signals.

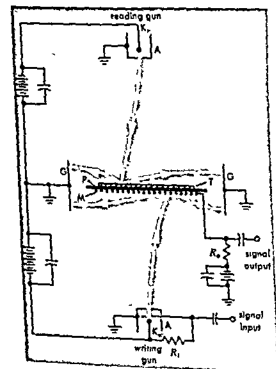


Fig 3 Signal-converter storage tube (Graphechon) (From M Knoll and B. Kazan, Storage Tubes, Wiley, 1952)

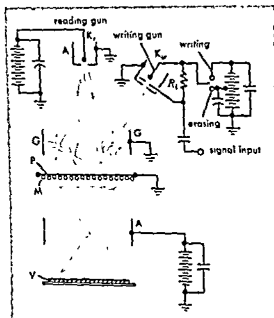


Fig 4 Viewing storage tube (From M Knoll and B. Kazan, Storage Tubes, Wiley, 1952)

Visual storage tubes. This group includes all electrostatic storage tubes that also provide a visual readout

Display storage tube This tube is closely akin to the recording storage tube. It is, however, of the electrical-visual family of storage tubes (Fig 4) In place of a scanned reading beam, the display storage tube employs a continuous, low-velocity flood beam from the reading gun, having a cross-sectional diameter equal to that of the storage surface In passing through the openings in the mesh M which supports the storage surface P, this beam is modulated point by point by the potential pattern established by writing. After passing the mesh plane, the flood beam (now modulated) is accelerated to a phosphor screen V where a visible display of the stored charge pattern is created

The display storage tube (variously known as the direct-view storage tube, the Iatron, or the Tono-

Storage tube

An electron tube into which information can be introduced and then extracted at a later time; also called a memory tube.

Operation. The process of introducing information into a storage tube is known as writing, and that of extracting useful information is called reading. The deliberate removal of information from the storage surface is called erasing. In some tubes reading automatically effects erasing; in others a separate operation is required. A charge-storage tube is one in which information is retained on a surface in the form of electric charges.

The characteristics of a storage tube are largely governed by the nature of the storage surface. This surface is usually a deposit or sheet of insulating or semiconducting material. The point-by-point potential of this surface is varied, in a controlled way, by the dielectric charging processes associated with electron bombardment of an insulator or semiconductor. The polarity of the resulting potential pattern depends upon the beam-accelerating voltage employed and the voltages applied to the elements adjacent to the storage surface. The low conductivity of the storage-surface material ensures that the charge (or potential) pattern will not be dissipated before the reading operation is initiated and completed.

In the reading operation, the potential pattern established by writing controls either the percentage of the primary beam that is able to pass through the openings of a fine mesh screen on which the storage surface may be deposited or else the magnitude of the secondary-electron current, which results when the primary beam strikes the storage surface. In either case, the current collected by the output electrode is modulated in accordance with the stored charge pattern.

Storage tubes may be classified according to the general nature of the input and output signals associated with them. Thus there are visual-electrical, electrical-electrical, and electrical-visual storage tubes. Visual-electrical storage tubes are better known as camera tubes and are described elsewhere (see TELEVISION CAMERA TUBE). In the following section several commercially important cathode-ray charge-storage tubes in the latter two classes will be briefly discussed.

Ordinary cathode-ray tubes have been used as electrical-electrical storage tubes in the buffer stages of high-speed digital computers. The phosphor screen and the envelope faceplate of these tubes serve together as the storage medium. An output electrode is provided by affixing a conducting membrane to the outside of the faceplate. Cathode-ray tubes especially designed for computer service are characterized by phosphor screens chosen for secondary-emission properties, stability, and purity, rather than for luminescent efficiency. These tubes are popularly called Williams tubes, after F. C. Williams, a pioneer in their application.

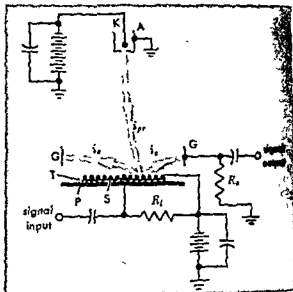


Fig. 1. Signal-converter storage tube (Radechon) (From M. Knoll and B. Kazan, Storage Tubes, Wiley, 1952)

Radechon. This vacuum tube, known also as a barrier-grid storage tube, is a single-gun device. The dielectric storage medium in this case is usually, but not necessarily, a sheet of mica T sandwiched between a continuous metal backing plate P and a fine mesh screen S (see Fig. 1). The electron beam i_{pr} is formed by the electron gun represented by cathode K and anode A. The collector G collects reading current i_c. This current develops an output signal voltage across output resistor R_o. The screen, the barrier grid, serves to restrict undesirable redistribution of secondary electrons over the storage surface. Additionally, the potential applied to the barrier grid is that to which the storage surface charges under the influence of beam bombardment. Writing consists of establishing localized voltage gradients, which may be of either polarity, across the storage layer. In the tube shown in Fig. 1 these gradients are established by varying the backplate potential by means of a signal fed into the input resistor R_l. The output current in reading is that which flows to the collector electrode when discharging this gradient.

The Radechon is highly versatile, finding use in simple delay schemes, signal-to-noise improvement applications, signal-comparison service, and systems for the conversion of signal-time bases.

Recording storage tube. This tube employs a storage medium in the form of an insulating deposit T on one side of a fine mesh screen S (Fig. 2). In the reading mode, the scanning electron beam approaches the storage mesh screen at low velocity. The magnitude of the reading beam current which is able to penetrate the openings of the storage mesh screen and impinge on the output electrode is controlled by the potential pattern established on the insulating deposit. This pattern is related to the input signal which modulates the electron beam during operation of the tube in the

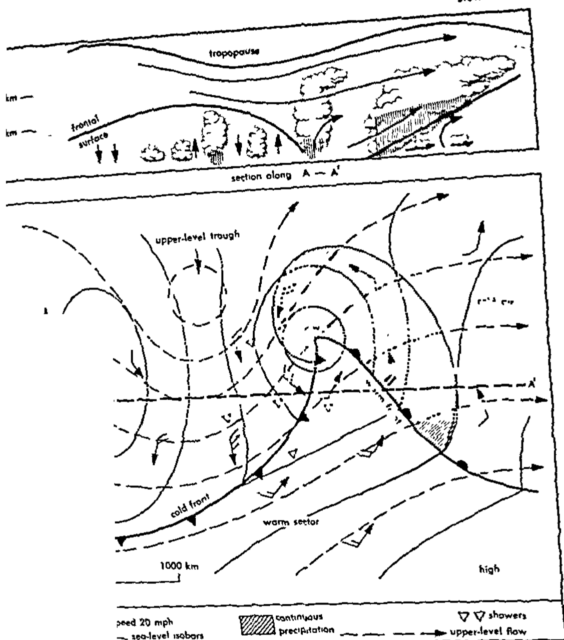


Fig 1 Schematic diagram of sea-level fronts and isobars in a wave cyclone, with lines of flow in upper troposphere superimposed (Mainly after Bjerkness and Palmén)

tions, with widespread cloud and precipitation but usually no pronounced concentration of stormy conditions Extensive cloudiness is also often present in the warm sector

Passage of the cold front is marked by a sudden wind shift, often with the onset of gusty conditions, with a pronounced tendency for clearing because of general subsidence behind the front Showers may be present in the cold air if it is [illegible]

Middle-latitude highs or anticyclones Extratropical cyclones alternate with high-pressure systems or anticyclones, whose circulation is generally opposite to that of the cyclone. The circulations of highs are not so intense as in well-developed cyclones, and winds are weak near their centers. In low levels, the air spirals outward from a high; descent in upper levels results in warming and drying aloft.

Anticyclones fall into two main categories, the warm "subtropical" and the cold "polar" highs The large and deep subtropical highs, centered over the oceans in latitudes 25–40° and separating the easterly trade winds from the westerlies of middle latitudes, are highly persistent

tron) is useful in many types of signal-display service, particularly in radar, sonar, and narrow-bandwidth television applications. This tube features high output brightness and a display persistence that is electrically controllable over a wide range, from milliseconds to many minutes.

Memotron. This is another electrical-visual storage tube which shares in general the operating principles and construction of the display storage tube. By choice of different storage-layer material, as well as slightly different electrode arrangements, this device sacrifices continuous tone-display capabilities for bistable visual-signal display, controllable in duration from a few milliseconds to infinity. This tube is suited to specialized oscillography.

Typotron. This tube is a variation of the Memotron. It substitutes a special writing gun for the conventional cathode-ray-type gun usually employed. This special gun includes a mask containing characters, such as the alphabet. The electron beam can be scanned over the mask to yield a resultant beam having a cross section resembling a chosen character. Such an electron gun is called a Charactron gun. The chosen characters are stored by the Memotron principle, providing a visually readable display. *See* CATHODE-RAY TUBE; DIGITAL COMPUTER [M.D.H.]

Storm

An atmospheric disturbance involving perturbations of the prevailing pressure and wind fields, including extratropical and tropical cyclones, blizzards, tornadoes and waterspouts, squalls, dust or sand storms, and thunderstorms.

Extratropical cyclones. These cyclonic disturbances are observed in all latitudes beyond about 20°, but they are most frequent in latitudes 30–60°. All extratropical cyclones of appreciable intensity form on or near the fronts separating the large air masses (*see* FRONT). The energy for their circulation is derived mainly from potential energy released by sinking of the colder (heavy) and rising of the warmer (light) air masses. This energy, which is augmented by latent heat released by condensation of water vapor, is converted into kinetic energy partly of the cyclonic disturbances and partly of the strong hemispheric westerlies in upper levels. In turn, the cyclones draw part of their energy from the upper wind systems.

The structure of a typical "wave" cyclone is shown in Fig. 1. In the Northern Hemisphere, winds blow about the low-pressure center in a counterclockwise fashion, nearly parallel to the isobars (lines of equal pressure). Wind strength is nearly proportional to the horizontal gradient of pressure.

Most cyclones develop from small disturbances on the polar front. After the wind circulation is initiated, the amplitude of the frontal wave—as viewed in the ground plan on a weather map—grows. At the stage shown in Fig. 1, the cyclone has a "warm sector" where relatively warm air is found at all levels. The frontal surface slopes up over the cold air masses poleward from the surface position shown, and warm air aloft extends in a tongue of decreasing depth northward past the cyclone center. Many cyclones progress to a more mature stage, in which the cold front gradually overtakes the warm front and forms an "occlusion" in which the warm air for some distance south of the low center is entirely lifted above the earth's surface.

Dynamical processes. The atmosphere is characterized by regions of horizontal convergence and divergence in which there is a net horizontal inflow or outflow of air in a given layer. Regions of appreciable convergence in the lower troposphere are always overlain by regions of divergence in the upper troposphere. As a requirement of mass conservation and the relative incompressibility of the air, low-level convergence is associated with rising motions in the middle troposphere, and low-level divergence with descending motions (subsidence).

In ascending the air cools by expansion and approaches saturation, while in descending the air becomes warmer and dryer. A central problem in forecasting cyclones and their accompanying weather patterns is predicting the distribution of convergence and divergence, from which the vertical motions are inferred. The convergence-divergence patterns can be deduced from the configurations and time changes of the pressure field by use of equations of motion relating it to the wind field.

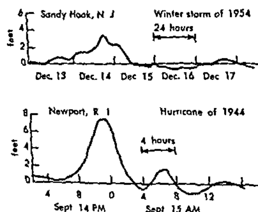
Viewed in an absolute frame of reference, the atmosphere is almost everywhere characterized by circulation in a cyclonic sense about a vertical axis. This circulation is due to rotation of the earth, augmented or decreased by rotation of the wind systems themselves. The absolute angular momentum of a given ring of air tends to be conserved, and contraction of the ring results in increasing the wind speed in a cyclonic sense. Thus low-level convergence is essential to establish and maintain the rotation in a cyclone. Friction at the earth's surface tends to retard the circulation.

Convergence in the surface cyclone is strongest on the east side and around the low center, while divergence is present on the west side behind the cold front. Compensation for the low-level convergence is provided by upper-level divergence, which is found on the east side of upper-level troughs. Cyclone formation and maintenance thus requires the presence of such a trough west of the cyclone as in Fig. 1. Divergence in such upper-level troughs is strongest near the belt of maximum upper-level westerly winds, and surface cyclones show a strong preference for the jet stream region, near the tropopause, where winds of 100–250 mph are observed.

Frontal storms and weather. Weather patterns in cyclones are highly variable, depending on moisture content and thermodynamic stability of air masses drawn into their circulations. Warm and occluded fronts, east of and extending into the cyclone center, are regions of gradual upgliding mo-

Storm surge

A transient, localized disturbance in sea level, resulting from the action of a tropical cyclone, an extratropical cyclone, or a squall over the sea. Storm surges, or storm tides, are not to be confused with tsunamis, or tidal waves, which result from seismic or molar disturbances of the earth. In the Northern Hemisphere those coastal regions which are particularly vunerable to storm surges include the periphery of the Gulf of Mexico, the Atlantic coast of the United States, the Gulf of Bengal, Japan and other islands of the Western Pacific which lie in the typhoon belt, and the coastal regions of the North Sea. The surges occurring in the North Sea originate from the actions of large-scale extratropical storms, particularly winter storms. On the east coast of the United States, hurricane-induced surges occur as well as surges originating from intense winter storms. In the Great Lakes and the Gulf of Mexico, surges resulting from squalls are known to occur; however, the hurricane-induced surges pose a more serious threat to the low-lying coastal areas of the Gulf. *See* TSUNAMI.



Examples of surge hydrographs

The time history of the surge at a given location at shore is represented by the surge hydrograph. This is a time sequence of the difference between the measured tide and the predicted periodic tide (see illustration). Maximum surge elevations of 15 ft above predicted tide are not uncommon. In the case of hurricane-induced surges, the peak water level seems to depend primarily upon the atmospheric pressure at the hurricane center. However, the horizontal scale, the direction and speed of propagation of the hurricane and the coastal geometry and bottom topography are important influencing factors in the storm surge behavior. When a hurricane crosses the coast from the sea, the greatest surge along shore usually occurs to the right of the hurricane path.

A storm surge is essentially a forced inertio-gravitational wave of great wave length. This implies that the duration or speed of the storm determines the dynamic augment of the water level at shore above that which would occur if the storm were stationary. Furthermore, the inertial character of the surges can explain the quasi-periodic resurgences which often follow the primary forced surge. *See* TIDE. [R.O.R.]

Bibliography: J. C. Freeman, Jr., L. Baer, and G. H. Jung, The bathystrophic storm tide, *J. Marine Research*, 16(1):12–22, 1957; D. Lee Harris, The hurricane surge, *Proc. 6th Conf. Coastal Eng.*, Council on Wave Research, pp. 96–114, 1958; A. C. Redfield et al., *Interaction of Sea and Atmosphere*, Meteorol. Monographs, 2(10), 1957.

Straight line mechanism

A linkage so proportioned and constrained that some point on the linkage describes a straight line, or nearly a straight line. The straight line mechanism, also called parallel motion mechanism, is seldom used to generate a straight line, having been replaced in most instances by a sliding block confined to a straight groove, or straight ways. The mechanical engine indicator, however, employs a straight line mechanism, and resourceful designers frequently use modifications of the many straight line mechanisms that have been proposed.

James Watt (1736–1819) devised the first straight line mechanism in 1784 and applied it to his vertical cylinder beam engine. Until that time, the lower end of a piston rod was guided by the piston in the cylinder, while the upper end was fastened to a chain that was wrapped on a sector which was fixed to the end of the walking beam and had its center at the pivot point of the beam. Watt's straight line motion made possible a double acting engine in which work could be done during the rising as well as the descending stroke of the piston, and supplied a positive means of guiding the outboard end of a piston rod. Thus it was an important contribution to the development of the steam engine since the crosshead and guide, which is used today (*see* SLIDER CRANK MECHANISM), did not appear until the 1820s.

Watt's straight line mechanism is shown in Fig. 1. Point *P* of the mechanism traces an approximate straight line only between points *a* and *b*.

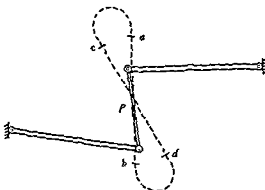


Fig. 1. Watt's straight line mechanism.

Cold anticyclones, forming in the source regions of polar and arctic air masses, decrease in intensity with height. Such highs may remain over the region of formation for long periods, with spurts of cold air and minor highs splitting off the main mass, behind each cyclone passing by to the south. Following passage of an intense cyclone in middle latitudes, the main body of the polar high may move southward in a major cold outbreak.

Blizzards are characterized by cold temperatures and blowing snow picked up from the ground by high winds. Blizzards are normally found in the region of strong pressure gradient between a well-developed arctic high and an intense cyclone. True blizzards are common only in the central plains of North America and Siberia and in Antarctica.

The principal cyclone tracks. Principal tracks for all cyclones of the Northern Hemisphere are shown in Fig 2 In middle latitudes, cyclones form most frequently off the continental east coasts and east of the Rocky Mountains.

Movements of cyclones, both extratropical and tropical, are governed by the large-scale hemispheric wave patterns in the upper troposphere The character of these waves is in part reflected by large circulation systems such as the subtropical highs, and the greatest anomalies from the principal cyclone tracks occur when these highs are displaced from the mean positions shown. Warm highs occasionally extend into high latitudes, blocking the eastward progression suggested by the average tracks and causing cyclones to move rather from north or south around their fringes.



Fig. 2. Principal tracks of extratropical cyclones and hurricanes with significantly associated features in the Northern Hemisphere.

Over the Mediterranean, cyclones form frequently in winter, but rarely in summer, when this area is occupied by an extension of the Saharan and Atlantic highs. Both the subtropical highs and the cyclone tracks in middle latitudes shift northward during the warmer months; on west coasts cyclones are infrequent or absent in summer south of latitudes 40–45°.

Tropical storms and disturbances. Weather in the tropics is influenced by migratory disturbances, including convergence patterns, vortices, and easterly waves. Land-sea distribution, orography, and diurnal variation of insolation modify the disturbance effects strongly, at any given place.

The intertropical convergence (ITC). A seasonally shifting zone of low pressure and variable winds, located between the northeast trade winds of the northern tropical areas and the southeast trades of the Southern Hemisphere, somewhat girdles the earth's low latitudes. Figure 2 shows its mean position during the northern summer. In Northern Hemisphere winter, the ITC is near 10–20°S from Africa east to the 180th meridian, and at about 5°N between 140°W and 20°E Movement of the ITC northward over Southeast Asia in spring marks the onset of the rainy summer monsoon

Horizontal convergence with uplifting of moist air masses in this ITC zone causes it to be one of the great rain belts, with frequent showers, thunderstorms, and squalls. The weather is highly variable, and at any one locale most of the seasonal rainfall is furnished by a few days' heavy rain

Tropical lows With concentrated bad weather on and near the ITC, shallow or weakly developed lows are generally found moving westward These do not attain appreciable pressure falls while near the equator, and only develop complete cyclonic circulations when removed poleward more than 5°. A minority of these develop into severe tropical hurricanes *See* HURRICANE.

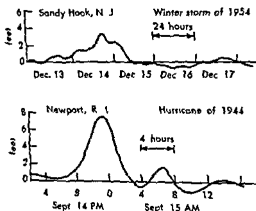
Easterly waves These are north-south-oriented low-pressure troughs in the general easterly current of the trade winds. Most such waves appear near latitude 20°, and in northern hemisphere terms, almost entirely between late spring and fall This is when the subtropical highs are farthest north and the easterly currents on their south sides are deep and extensive

On approach of an easterly wave, the wind backs to a northeasterly direction; on passage of the trough the wind veers to southeast and then gradually returns to easterly Good weather prevails west of the trough, with frequent moderate to heavy showers and thunderstorms on its east side The waves, moving westward with speeds around 15 mph, pass at intervals of 3–4 days, and showery weather may last for 2 days after trough passage [C.W.N.]

Bibliography: S. Petterssen, *Weather Analysis and Forecasting*, vol. 1, 1956; H. Riehl, *Tropical Meteorology*, 1954

Storm surge

A transient, localized disturbance in sea level, resulting from the action of a tropical cyclone, an extratropical cyclone, or a squall over the sea. Storm surges, or storm tides, are not to be confused with tsunamis, or tidal waves, which result from seismic or molar disturbances of the earth. In the Northern Hemisphere those coastal regions which are particularly vulnerable to storm surges include the periphery of the Gulf of Mexico, the Atlantic coast of the United States, the Gulf of Bengal, Japan and other islands of the Western Pacific which lie in the typhoon belt, and the coastal regions of the North Sea. The surges occurring in the North Sea originate from the actions of large-scale extratropical storms, particularly winter storms. On the east coast of the United States, hurricane-induced surges occur as well as surges originating from intense winter storms. In the Great Lakes and the Gulf of Mexico, surges resulting from squalls are known to occur; however, the hurricane-induced surges pose a more serious threat to the low-lying coastal areas of the Gulf. *See* TSUNAMI.



Examples of surge hydrographs

The time history of the surge at a given location at shore is represented by the surge hydrograph. This is a time sequence of the difference between the measured tide and the predicted periodic tide (see illustration). Maximum surge elevations of 15 ft above predicted tide are not uncommon. In the case of hurricane-induced surges, the peak water level seems to depend primarily upon the atmospheric pressure at the hurricane center. However, the horizontal scale, the direction and speed of propagation of the hurricane, and the coastal geometry and bottom topography are important influencing factors in the storm surge behavior. When a hurricane crosses the coast from the sea the greatest surge along shore usually occurs to the right of the hurricane path.

A storm surge is essentially a forced inertio-gravitational wave of great wave length. This implies that the duration or speed of the storm determines the dynamic augment of the water level at shore above that which would occur if the storm were stationary. Furthermore, the inertial character of the surges can explain the quasi-periodic resurgences which often follow the primary forced surge. *See* TIDE. [R.O.R.]

Bibliography: J. C. Freeman, Jr., L. Baer, and G. H. Jung, The bathystrophic storm tide, *J. Marine Research*, 16(1):12–22, 1957; D. Lee Harris, The hurricane surge, *Proc. 6th Conf. Coastal Eng.*, Council on Wave Research, pp. 96–114, 1958; A. C. Redfield et al., *Interaction of Sea and Atmosphere*, Meteorol. Monographs, 2(10), 1957.

Straight line mechanism

A linkage so proportioned and constrained that some point on the linkage describes a straight line, [illegible] a sliding block confined to a straight groove, or straight ways. The mechanical engine indicator, however, employs a straight line mechanism, and resourceful designers frequently use modifications of the many straight line mechanisms that have been proposed.

James Watt (1736–1819) devised the first straight line mechanism in 1781 and applied it to his vertical cylinder beam engine. Until that time, the lower end of a piston rod was guided by the piston in the cylinder, while the upper end was fastened to a chain that was wrapped on a sector which was fixed to the end of the walking beam and had its center at the pivot point of the beam. Watt's straight line motion made possible a double acting engine in which work could be done during the rising as well as the descending stroke of the piston, and supplied a positive means of guiding the outboard end of a piston rod. Thus it was an important contribution to the development of the steam engine since the crosshead and guide, which is used today (*see* SLIDER CRANK MECHANISM), did not appear until the 1820s.

Watt's straight line mechanism is shown in Fig. 1. Point *P* of the mechanism traces an approximate straight line only between points *a* and *b*.

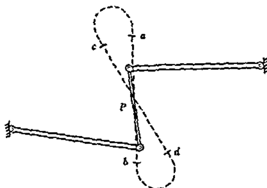


Fig. 1 Watt's straight line mechanism

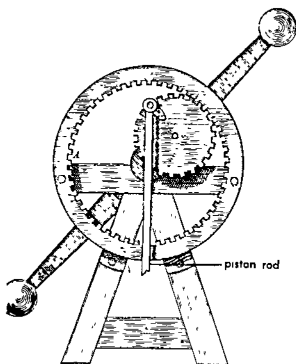


Fig. 2. Epicyclic straight line mechanism (*Smithsonian Institution*)

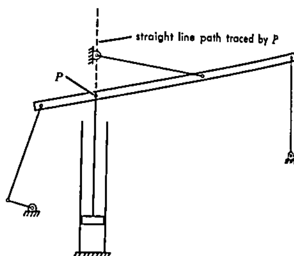


Fig. 3. Grasshopper linkage.

An early departure from Watt's mechanism is shown in Fig. 2. The end of the piston rod is pivoted on the pitch circle of the smaller gear of the epicyclic train. Although this epicyclic mechanism was considered impractical at the time of its invention because of imperfection of the gears, point *a* of the linkage does describe a true straight line.

The grasshopper linkage shown in Fig. 3 was used in an engine built by Oliver Evans (1755–1819) and in a somewhat modified form in George Stephenson's Stourbridge Lion, the first locomotive brought to America. The stationary engine that employed this linkage was called a grasshopper engine, because of the action of the linkage.

Richard Roberts (1789–1864), who built the first metal planer (1817) which provided a practical means of making straight metal guides for a slider, contributed also to the catalog of straight line mechanisms. Roberts' linkage is shown in Fig. 4*a*.

The straight line mechanism of Fig. 4*b* was attributed by William J. M. Rankine to John Scott Russell (1808–1882), an English engineer. This modification of the grasshopper linkage traces an exact straight line if the slider is confined to a straight path, but it has the disadvantage of requiring a sliding block and guide.

The first exact straight line motion to employ turning pairs only (Fig. 4*c*) was proposed about 1864 by C. N. Peaucellier (1832–1913). It has been little used because of its complexity and because it came when the acute need for such a device was past. [E S F.]

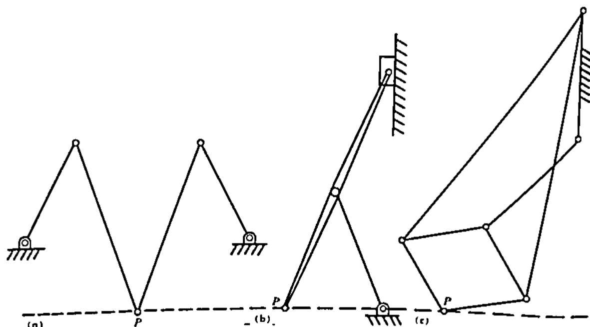


Fig 4 Straight line mechanisms. (*a*) Roberts' mechanism (*b*) Russell's linkage (*c*) Peaucellier mechanism.

Strain

Deformation or change in shape of a material as a consequence of applied forces. Strain is directly measurable (*see* STRAIN GAGE), and from such measurement, within the elastic limit, the internal stress that accompanied the strain can be determined (*see* ELASTIC LIMIT; HOOKE'S LAW; PHOTOELASTICITY, YOUNG'S MODULUS). The strain-producing action sets up consequent stresses in the material, and stresses in a material cause a deformation or strain. That is, an initial strain is always accompanied by a stress. For this reason, the two phenomena are usually dealt with together (*see* STRESS AND STRAIN). [F.H R]

Strain gage

A device which uses the change of electrical resistance of a wire under tension to measure pressure

The strain gage converts a mechanical motion to an electrical signal by use of the fact that when a wire is stretched, its length is increased and its diameter is decreased; its electrical resistance is therefore increased The change in resistance is a measure of the mechanical motion, which in turn is a measure of the pressure *See* PRESSURE TRANSDUCER

The complete pressure-measuring instrument system comprises a sensing element usually a bourdon tube, bellows, or diaphragm element, the strain gage (bonded or unbonded); and an indicating or recording instrument

The bonded strain gage is affixed to a piece of metal (such as a bourdon tube) which deforms elastically as pressure is applied The usual form of the bonded strain gage is a flat zigzag grid of very fine (perhaps 0.001-in diameter) wire imbedded in a thin sheet of impregnated paper. The sheet is cemented to the pressure-sensing element

Bonded strain gages are available in a wide range of resistances (50–5000 ohms), of sizes (1/16–6 in), and of materials, to meet the requirements of various applications

Unfortunately most wires which have desirable characteristics as strain gage material are also sensitive to temperature (change resistance when the temperature changes) This effect is not always significant in dynamic testing but for tests of long duration or for continuous processing, the effect of temperature must be minimized The circuit of Fig 3 minimizes the temperature effect with a temperature compensator identical to the measuring gage mounted on an unstressed element at the same temperature as the measuring gage

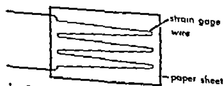


Fig 1. Bonded strain gage

The unbonded strain gage consists of a grid of fine wires strung under slight tension between a stationary frame and a movable armature. Pressure applied to the bellows or diaphragm sensing element moves the armature with respect to the frame, increasing tension in one-half of the filaments and decreasing the tension in the rest. Proper arrangement of the wires in a Wheatstone bridge circuit will nearly nullify the effect of temperature

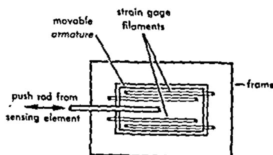


Fig 2 Unbonded strain gage.

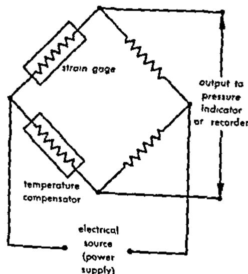


Fig 3 Simple strain-gage circuit

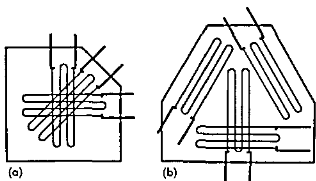
A strain gage is almost always used in a Wheatstone bridge circuit When pressure is applied, the resistance of the strain gage changes, and the output voltage of the bridge changes. The output voltage cannot be measured by a conventional voltmeter, and so common practice—especially in industry—is to amplify the output to drive an indicating or recording instrument with a self-balancing bridge.

For high-speed dynamic measurements, a moving-mirror oscillograph can furnish response as fast as 2000 cps, and an oscilloscope (using a cathode-ray tube) can give even faster response.

Strain-gage accuracies may be from 0 1 to 2% of full scale, depending on materials and design For example, instruments used for dynamic measurements may have excessive drift. See PRESSURE MEASUREMENT. [B D H.; H G P.]

Strain rosette

A pattern of intersecting gage lines on a surface along which linear strains are measured to find stresses at a point. A rosette gage is an assembly of strain measuring components arranged to measure strains in the directions of the respective gage lines. To facilitate computations, three lines are usually oriented to form a rectangular or 45° rosette, or as an equiangular rosette, also called the delta or 60° rosette as illustrated.



Strain rosettes (a) 45° type; (b) 60° type.

Electrical resistance-type gages are usually employed (*see* STRAIN GAGE). With suitable instrumentation, strains are recorded in microinches per inch. Strains in the selected directions can also be measured by mechanical gages The three measured strains permit evaluation of strains in any direction. The maximum and minimum strains, called principal strains can be found analytically or by Mohr's circle of strain, and the corresponding stresses by the generalized Hooke's law. This procedure for finding the maximum stresses and their directions from linear strain measurements is called the rosette method. *See* STRESS AND STRAIN

[W J KR]

Bibliography: M Hetenyi (ed), *Handbook of Experimental Stress Analysis*, 1950.

Strand line

A line at the margin or shore of a sea or lake In geology, strand lines of ancient seas can be identified by recognizing sedimentary structures and textures, and the organisms that characterize shores It is difficult to recognize shores of identical time. Seas spread and retreat in time, causing shore facies to migrate laterally through rock sequences. Organisms are more sensitive to environments than to small time differences, so planes of synchroneity are more difficult to determine than sites of deposition. Changes in sea level (eustatic movements) are universal, so their effects on strand lines have been applied in making world-wide correlations in time. *See* FACIES (GEOLOGY); SEA LEVEL FLUCTUATIONS; WARPING, EARTH CRUST.

[M K.]

Bibliography: *Finding Ancient Shorelines*, Society of Economic Paleontologists and Mineralogists Spec. Publ 3, 1955.

Strange particle

An inclusive name for K-mesons and hyperons. These particles were initially considered strange because they have relatively long lifetimes (in the range 10^{-8}–10^{-10} sec), which appeared inconsistent with their copious production in high energy nuclear collisions (lifetimes of about 10^{-23} sec were expected for K-mesons and hyperons with strong nuclear interactions). This inconsistency was resolved by the recognition of a new quantum number, the strangeness number s, together with a new physical law (conservation of strangeness) which these nuclear processes satisfy but which these decay processes violate; for a further statement on the role of s and the values assigned to various particles, *see* MESON. The known strange particles are distinguished by having a nonzero value of s; for discussion of their production processes, reactions, and decay properties, *see* HYPERON; MESON *See also* SELECTION RULES (PHYSICS).

[R H D]

Stratigraphic nomenclature

A system of naming used by geologists in classifying the rock record of the geologic history of the earth Sedimentary rocks, laid down layer by layer in seas, lakes, river floodplains, or elsewhere, are the principal record of that history The record is complex, for rocks vary greatly not only vertically from layer to layer but also, though over greater distances, horizontally along the layers.

From the first it was realized that any succession of such layers represents a succession in time, the lower layers being the older (law of superposition, Steno, 1669). Accordingly, stratigraphic nomenclature has always attempted to express the time relations of the various layers as well as their intrinsic physical character Extension of such time relations laterally from one succession to another is called correlation (*see* STRATIGRAPHY) Within a local area the rocks can be correlated by tracing individual layers or groups of layers, but for great distances, especially from continent to continent, correlation depends on the proved generalization that the fossils in the layers of a given age differ from those in layers of all other ages (law of faunal succession, W Smith, 1799; explained by theory of organic evolution, Darwin, 1859)

Categories of stratigraphic subdivisions. Schemes of stratigraphic nomenclature grew up rather haphazardly through the nineteenth century, but attempts to standardize them began late in that century At present, two main schemes are prevalent—a North American and a European—though considerable divergences remain in the usage on each continent. In North America, a code of stratigraphic nomenclature was published in 1933, and since 1946 the American Commission on Stratigraphic Nomenclature has been actively working to bring about agreement on principles as well as uniformity in practice. No comparable body exists in Europe; a commission in Australia has pub-

lished a code agreeing in essentials with American usage. Since 1952, the International Commission on Stratigraphy has been preparing lexicons of the stratigraphic terms used on the various continents. It is also trying to formulate grounds for international agreement on principles and categories of units.

Continental European usage emphasizes time (as recorded by fossil content) as the chief basis for stratigraphic nomenclature. The stratigraphic subdivisions recognized are intended to embrace all layers laid down during a given time interval; in practice they include all layers containing the fossils of one fauna in the succession of fossil faunas. Table 1 gives a hierarchy of such subdivisions. Stratigraphic units based directly on the physical character of the rocks without reference to time are recognized in European usage, but they are considered informal or only preliminary in nature.

Table 1 Table of stratigraphic units

Geologic-time units	Time-stratigraphic units	Rock-stratigraphic units*
Era*†	—	(Group)†
Period*†	System*†	(Formation)†
Epoch*†	Series*†	(Member)†
Age*†	Stage*†	(Bed)†
Phase*	Zone*†	

* Continental European usage—scheme of stratigraphic units accepted by the Eighth International Geological Congress in 1900.
† [illegible]

North American usage, on the other hand, considers that units based directly on the characters of the rocks—rock-stratigraphic units—should have a separate formal status equal to that of the time-stratigraphic units (a dissenting minority would simply range them as lower members of the time-stratigraphic hierarchy). The rock-stratigraphic units are also listed in Table 1, but they should not be considered parallel to the other kinds of units in that table (as those kinds of units are to each other). Since 1950 the American Commission has sponsored a trend in North American usage [illegible] and related [illegible] are placed in the latter category and stage, series, and system in the former. Dissent[illegible] zone [illegible] based on fossil content as [illegible]

Usage in Australia and in most major oil-producing areas tends to agree with North American usage. There also are advocates of this usage in Europe, especially in Great Britain and Scandinavia. Soviet usage, codified in 1956, is based on the European tradition, though with a few changes in terms. It recognizes the need for local units (series, suite, bundle) in addition to the units of the main time-stratigraphic scale. It assigns them only an auxiliary place, however, and the largest local unit used in any area is subordinated to the smallest unit of the main scale there used.

Names of stratigraphic subdivisions. The stratigraphic nomenclatures discussed above were worked out in and for the generally fossiliferous rocks deposited during the last 500,000,000 years (the Phanerozoic eon, the time since the appearance of the olenellid-archeocyathid fauna at the beginning of the Cambrian period). The rocks deposited in earlier times (Cryptozoic eon or Precambrian time) contain no fossils that can be used in correlation. Attempts to classify those rocks into time-stratigraphic units, though numerous, have been mutually contradictory. Methods have been developed since the 1920s for determining rock ages in years by measuring radioactive minerals for parent and daughter elements. *See* RADIOACTIVE MINERALS; ROCK (AGE DETERMINATION). The scarcity of suitable minerals in unaltered sedimentary rocks means, however, that ages so determined generally indicate the dates of igneous intrusions or of widespread metamorphism rather than of original sedimentary deposition; the methods have errors of 5 per cent or more.

Table 2 gives the subdivision of the Phanerozoic eon into eras and periods as now accepted almost throughout the world (the *principal divergences are noted*).

Table 2 Internationally accepted subdivisions of geologic time (youngest at top)

Eras	Periods		Epochs
Cenozoic	Quaternary (era*)		Holocene (Recent)
			Pleistocene
	Tertiary (era*)	Neogene*	Pliocene
			Miocene
		Paleogene*	Oligocene
			Eocene
			Paleocene
Mesozoic	Cretaceous		
	Jurassic		
	Triassic		
Paleozoic	Permian		
	Carboniferous		Pennsylvanian†
			Mississippian†
	Devonian		
	Silurian	Silurian‡	Gothlandian‡
	Ordovician		Ordovician‡
	Cambrian		

* Current French usage.
† Current North American usage.
‡ Current French and German usage.

Subdivisions of the periods or systems into epochs or series is less uniform. Table 2 also gives a common subdivision of the Cenozoic era into epochs, but Paleocene and especially Holocene have met with *much* opposition. Some of the older

periods (Jurassic, Triassic, Devonian, Cambrian) are almost universally divided into three subdivisions, but for others (Cretaceous, Carboniferous) a division into two parts is preferred, or no agreement has been reached Some of the epochs have their own special names; others are simply designated Late, Middle, and Early (the series are Upper, Middle, and Lower).

Stages and zones are in general much more provincial, though in a few systems, notably the Jurassic and Cretaceous, the stages established in western Europe have been recognized practically around the world Stages mostly bear the names of places, commonly with the ending "-ian" (Moscovian, Oxfordian, Delmontian stages); each zone bears the name of a characteristic fossil (the zones of *Pseudoschwagerina*, of *Cardioceras cordatum*, of *Hyracotherium*).

The rock-stratigraphic units of the North American scheme are defined as lithogenetic units, that is, units formed under essentially uniform (or uniformly alternating) conditions. As their principal purpose is to serve as units for detailed geologic mapping or local description (as in studies of cuttings or cores recovered from oil wells), objectivity is of prime importance in their designation. They are named for geographic localities near which they are typically exposed; such type sections play much the same role that type specimens play for the units of biological nomenclature No two units in the same country are supposed to bear the same geographic name, and priority is generally accepted as a principle of nomenclature, though more exceptions are permitted than in biology The term for the principal rock type present may also be part of the name (Knox dolomite group, Austin chalk, Monterey formation). For groups and formations, formal naming according to this scheme is now considered obligatory, but for members it is optional, and the names of beds are considered outside formal stratigraphic nomenclature

The general principles of rock-stratigraphic nomenclature are also extended to igneous and metamorphic rocks. The units are in all ways comparable to rock-stratigraphic units in sedimentary rocks. Nonsedimentary rocks are also assigned to time-stratigraphic units (for metamorphic rocks, based on the time of their original formation, not their metamorphism). Except for the volcanic igneous rocks, however, it is rarely possible to assign them with any precision [J.R.]

Bibliography: C. O. Dunbar and John Rodgers, *Principles of Stratigraphy*, 1957; W. C. Krumbein and L. L. Sloss, *Stratigraphy and Sedimentation*, 1951.

Stratigraphy

The branch of the science of geology that studies layered or stratified rocks. Chiefly it concerns sedimentary rocks, but its principles may also be applied to layered igneous rocks, such as lavas and tuffs, and to metamorphic rocks that were formed from sedimentary or volcanic rocks. *It deals with* the observed interrelations of the layers of such rocks and with the historical conclusions that can be inferred from those interrelations. Other aspects of sedimentary geology are sedimentary petrography (the study of the materials composing sedimentary rocks) and sedimentation (the study of the processes by which sediments are formed at present). No sharp line can be drawn between these subjects, however, and each depends in part on the conclusions of the others. *See* SEDIMENTARY ROCKS; SEDIMENTATION (GEOLOGY).

Objectives of stratigraphy. The first task of stratigraphy is the description of local sequences of strata; from these descriptions local geologic history can be inferred by using the law of superposition, which states that in a local sequence of rock layers, the lower ones are the older. First deduced by N. Steno (1669), the law is amply established by studies of sedimentation

The second task of stratigraphy is the correlation of these local sequences, that is, the determination of their mutual time relations and the integration of the local histories into a regional or world wide chronologic framework. Correlation can be accomplished in several different ways, but historically much the most important has been by fossils, using the law of faunal succession, which states that rocks with the same fossil fauna or flora in different parts of the world are of roughly the same age (the converse is only partially true). First empirically worked out by W. Smith (1799), this law has been verified in all parts of the world and is now explained by the theory of organic evolution as the expression of the gradual development of organic life on earth (Darwin, 1859). *See* EVOLUTION, ORGANIC.

The third task of stratigraphy is interpretation of the geologic history of the earth from the scattered data of local sequences and criteria of correlation The conclusions of sedimentation, sedimentary petrology, and plant and animal ecology are all called on to provide clues for this interpretation, in accordance with the principle of uniformity. The data of other branches of geology are also pertinent These include structural geology, geomorphology, and igneous and metamorphic petrology. This aspect of the subject merges with (and is often called) historical geology. *See* GEOLOGY.

Applications of stratigraphy. Stratigraphy is of great practical value. Many of its basic principles were first discovered by copper miners in central Germany and coal miners in Great Britain, and they are obviously applicable to bedded mineral deposits of all sorts (for example, coal, iron ore, and phosphate) Stratigraphy's greatest extension has come about, however, through its application to petroleum exploration Oil and natural gas occur almost exclusively in stratified rocks, the oil and gas are entrapped mainly where permeable strata are overlain and laterally surrounded by impermeable strata, the favorable configuration of strata being produced by original deposition differences, by later changes in porosity or permeability, *by later warping or breaking of the strata, or* by various combinations of th [illegible] ls drilled

up such trapped accumulations yield much information on the strata penetrated, and new methods for obtaining more information from them are constantly being devised. As a result, the stratigraphy of oil-bearing regions is no longer dependent on observations at the earth's surface but can be studied in three dimensions in considerable detail.

Theoretical conclusions. Several important general theoretical conclusions about the history of the earth can be drawn from stratigraphy; most of these are expressions of the principle of uniformity, of which stratigraphic succession is one of the best exemplifications. In general, the same conditions of sedimentation—marine and nonmarine—have prevailed throughout the decipherable stratigraphic record (probably roughly 3×10^9 years). At one time the oldest rocks were thought to display unique characters, but this view has now been fairly well discredited. As a corollary, liquid water has always been a major agent of erosion and sedimentation, and hence the average temperature of the earth's surface must have been within the range of liquid water over the same time span, and presumably the earth's climatic pattern in the past could not have been very different from what it is now.

Stratigraphy is also responsible for the major generalization that great mountain ranges coincide with former belts of exceptionally thick sedimentary rocks, deposited in long but relatively narrow linear troughs called geosynclines. The nature of the genetic relation between geosynclines and the subsequent mountain ranges is still controversial and is the subject of continuing investigation. Also for many years the record was believed to show that virtually all marine deposits now found on the continents, including those in geosynclines, were formed in shallow seas spreading over the continental blocks rather than in deep ocean basins or troughs and hence that continents and ocean basins are permanent features of the earth's crust, geosynclines being accidents within the continental masses. Studies during the 1950s suggest, however, that many sedimentary rocks [illegible] ran [illegible] tro[illegible] con[illegible] ...e of the present East and West Indies (but not in open ocean basins far from the continents). The further conclusion has been drawn by many that the continents have accreted spasmodically through geologic time by the incorporation of such island-arc–trough systems into the continent in the process of mountain building, followed by the development of new systems along the new continental borders. See GEOSYNCLINE, OROGENY, TECTONIC PATTERNS.

Finally, stratigraphy provides a record, though still somewhat incomplete, of the ever increasing complexity and deployment of life during the last 500 000 000 years of the earth's history, including some of the most impressive evidence, both in detail and in the large, for the theory of organic evolution. See PALEOBOTANY; PALEONTOLOGY.

[J.R.]

Stratosphere

Region of the atmosphere lying above the troposphere and tropopause with isothermal or very stable temperature gradients. Originally this was called the isothermal layer by Teisserenc de Bort, its discoverer. By convention, the upper boundary is often placed at 30 km where, in middle and high latitudes, the temperature gradient becomes more positive. The stratosphere is very dry compared with the troposphere, and clouds are extremely rare. High winds are present, but there may be little turbulent mixing. The seasonal temperature range increases rapidly with height. The classical theories of E. Gold and R. Emden assume that the stratosphere is in radiative equilibrium. See ATMOSPHERE.

[R.M.G.]

Strawberry

Low-growing perennials, spreading by stolons, with fruit consisting of a fleshy receptacle, and "seeds" in pits or nearly superficial on the receptacle. See FRUIT (BOTANY).

The strawberry, as known in the United States, is derived from two species, *Fragaria chiloensis*, which grows along the Pacific Coast of North and South America, and *F. virginiana*, the eastern meadow strawberry, both members of the plant order Rosales. These species were apparently crossed in Europe in the early eighteenth century, and some of the hybrid offspring brought back to North America. The European strawberry is a smaller-fruited type, usually grown from seed. See SEED (BOTANY).

Distribution and importance. The strawberry is the most universally grown of the small fruits, both in the home garden and in commercial plantings. See FRUIT GROWING (SMALL). Home garden production is possible in nearly all of the states provided water can be supplied where rainfall is insufficient. Commercial production is important in probably three-fourths of the states. Although the acreage may vary greatly from year to year, the following states are large producers: Oregon, California, Tennessee, Michigan, Louisiana, Washington, Arkansas, Kentucky, and New York. The crop is worth over $50,000 000 annually.





Strawberry. (a) Flower. (b) Fruit. (From L. H. Bailey, ed., The Standard Cyclopedia of Horticulture, vol. 3, Macmillan, 1935)

periods (Jurassic, Triassic, Devonian, Cambrian) are almost universally divided into three subdivisions, but for others (Cretaceous, Carboniferous) a division into two parts is preferred, or no agreement has been reached. Some of the epochs have their own special names; others are simply designated Late, Middle, and Early (the series are Upper, Middle, and Lower).

Stages and zones are in general much more provincial, though in a few systems, notably the Jurassic and Cretaceous, the stages established in western Europe have been recognized practically around the world. Stages mostly bear the names of places, commonly with the ending "-ian" (Moscovian, Oxfordian, Delmontian stages); each zone bears the name of a characteristic fossil (the zones of *Pseudoschwagerina*, of *Cardioceras cordatum*, of *Hyracotherium*).

The rock-stratigraphic units of the North American scheme are defined as lithogenetic units, that is, units formed under essentially uniform (or uniformly alternating) conditions. As their principal purpose is to serve as units for detailed geologic mapping or local description (as in studies of cuttings or cores recovered from oil wells), objectivity is of prime importance in their designation. They are named for geographic localities near which they are typically exposed; such type sections play much the same role that type specimens play for the units of biological nomenclature. No two units in the same country are supposed to bear the same geographic name, and priority is generally accepted as a principle of nomenclature, though more exceptions are permitted than in biology. The term for the principal rock type present may also be part of the name (Knox dolomite group, Austin chalk, Monterey formation). For groups and formations, formal naming according to this scheme is now considered obligatory, but for members it is optional, and the names of beds are considered outside formal stratigraphic nomenclature.

The general principles of rock-stratigraphic nomenclature are also extended to igneous and metamorphic rocks. The units are in all ways comparable to rock-stratigraphic units in sedimentary rocks. Nonsedimentary rocks are also assigned to time-stratigraphic units (for metamorphic rocks, based on the time of their original formation, not their metamorphism). Except for the volcanic igneous rocks, however, it is rarely possible to assign them with any precision [J.R.]

Bibliography: C. O. Dunbar and John Rodgers, *Principles of Stratigraphy*, 1957; W. C. Krumbein and L. L. Sloss, *Stratigraphy and Sedimentation*, 1951.

Stratigraphy

The branch of the science of geology that studies layered or stratified rocks. Chiefly it concerns sedimentary rocks, but its principles may also be applied to layered igneous rocks, such as lavas and tuffs, and to metamorphic rocks that were formed from sedimentary or volcanic rocks. It deals with the observed interrelations of the layers of such rocks and with the historical conclusions that can be inferred from those interrelations. Other aspects of sedimentary geology are sedimentary petrography (the study of the materials composing sedimentary rocks) and sedimentation (the study of the processes by which sediments are formed at present). No sharp line can be drawn between these subjects, however, and each depends in part on the conclusions of the others. *See* SEDIMENTARY ROCKS, SEDIMENTATION (GEOLOGY).

Objectives of stratigraphy. The first task of stratigraphy is the description of local sequences of strata; from these descriptions local geologic history can be inferred by using the law of superposition, which states that in a local sequence of rock layers, the lower ones are the older. First deduced by N. Steno (1669), the law is amply established by studies of sedimentation.

The second task of stratigraphy is the correlation of these local sequences, that is, the determination of their mutual time relations and the integration of the local histories into a regional or world wide chronologic framework. Correlation can be accomplished in several different ways, but historically much the most important has been by fossils, using the law of faunal succession, which states that rocks with the same fossil fauna or flora in different parts of the world are of roughly the same age (the converse is only partially true). First empirically worked out by W. Smith (1799), this law has been verified in all parts of the world and is now explained by the theory of organic evolution as the expression of the gradual development of organic life on earth (Darwin, 1859). *See* EVOLUTION, ORGANIC.

The third task of stratigraphy is interpretation of the geologic history of the earth from the scattered data of local sequences and criteria of correlation. The conclusions of sedimentation, sedimentary petrology, and plant and animal ecology are all called on to provide clues for this interpretation, in accordance with the principle of uniformity. The data of other branches of geology are also pertinent. These include structural geology, geomorphology, and igneous and metamorphic petrology. This aspect of the subject merges with (and is often called) historical geology. *See* GEOLOGY.

Applications of stratigraphy. *Stratigraphy* is of great practical value. Many of its basic principles were first discovered by copper miners in central Germany and coal miners in Great Britain, and they are obviously applicable to bedded mineral deposits of all sorts (for example, coal, iron ore, and phosphate). Stratigraphy's greatest extension has come about, however, through its application to petroleum exploration. Oil and natural gas occur almost exclusively in stratified rocks; the oil and gas are entrapped mainly where permeable strata are overlain and laterally surrounded by impermeable strata, the favorable configuration of strata being produced by original deposition differences, by later changes in porosity or permeability, by later warping or breaking of the strata, or by various combinations of [illegible] s drilled

ble in an earphone worn by the observer. The number of clicks in a clocked period of time can be translated into velocity by a calibration table.

Direct measurements of discharge by current meter are made periodically, usually once a month. To compute the flow during the intermediate period, it is necessary to observe the stage (elevation of water surface), which fluctuates in response to storms and ground water yield to the stream. This is done by a water-stage recorder, which makes a graphical record of water level in a well that is connected by an open pipe to the natural stream. Fluctuations of water surface are followed by a float connected by a wire to the recorder.

The stage record traced by pen on a clock-driven paper is translated into discharge by use of a graphical relation of stage to discharge established by the successive direct measurements made with current meter.

In 1958, about 7000 gaging stations on rivers in the United States and its territories were maintained by the U.S. Geological Survey. A much smaller number of stations is maintained by a few state, international, and other federal agencies. The earliest stream-flow records in the United States began about 1890.

A stream-flow record is a time sample of fluctuating river flow, and is subject to the same principles that govern the use of sampling techniques in other fields. Modern stream-gaging networks are based on the principle that a primary permanent net of selected stations will provide long term samples of time variations. Other roving or satellite stations are maintained long enough at a given place to provide a correlation with one of the primary stations. The roving station can then be moved to sample another location. After the correlation is established, estimates of flow characteristics sufficiently accurate for most purposes can be made by utilizing the record at a primary station in conjunction with the correlation graph. See HYDROLOGY, SURFACE WATER. [L.B.L.]

Stream transport and deposition

The sediment debris load of streams is a natural corollary to the degradation of the landscape by weathering and erosion. Eroded material reaches stream channels through rills and minor tributaries, being carried by the transporting power of running water and by mass movement, that is, by slippage, slides, or creep. The size represented may vary from clay to boulders. At any place in the stream system the material furnished from places upstream either is carried away, or, if there is insufficient transporting ability, is accumulated as a depositional feature. The accumulation of deposited debris tends toward increased ease of movement, and this tends eventually to bring into balance the transporting ability of the stream and the debris load to be transported.

Stream loads. Because streams form and adjust their own channels, the debris load to be carried and the ability to carry load tend to reach and maintain a quasi-equilibrium. A reach of stream (part of the course) which attains this equilibrium is considered graded.

Much has been written concerning the concept of the graded stream. At one time absence of waterfalls or other discontinuities of longitudinal profile was considered necessary and, in fact, evidence for the condition of grade. Because much remains to be learned about the mechanics of debris transportation, the criteria for the graded condition may be expected to be extended and revised. In the present state of knowledge, however, it appears acceptable to think of reaches or segments of channel being graded, even when separated by reaches not so adjusted. A graded stream is one in which, over a period of years, slope and channel characteristics are delicately adjusted to provide, with available discharge, the shear forces required for the transportation of the load supplied from the drainage basin.

Two terms which have been useful to geologists and engineers dealing with rivers are competence and capacity. Competence was used by G. K. Gilbert to mean the ability to move debris, and its measure is the maximum size of material which can barely be moved. Capacity of a stream is the total load which it can carry under given conditions and is measured as weight of debris moved per unit of time. The usefulness of these terms has lessened with demand for increasingly quantitative description of stream action. Direct measurement of the total load of a stream is not possible with present instrumentation. Sampling equipment now in general use measures only the suspended portion of the debris and not that moving close to the streambed. Thus, except in special situations, the carrying capacity of a stream cannot be precisely measured, and available theory allows only an approximation of total load by computation.

The maximum size of debris which can be carried varies, depending on subtle variations of several factors. Thus competence, a highly useful concept, cannot be determined with satisfaction either in the field or by computation. The concepts implied by these terms will gain even greater value and importance as both theory and field measurement techniques improve. The following review of the present status of theory of debris transport will [illegible] how the usefulness of these con- [illegible] ...ine quan- [illegible] ...n. Debris [illegible] ... hydro- [illegible] ...own [illegible] fluid alters in an important [illegible] draulic relationships which applied to a fixed bed. For example, in a movable-bed channel, boundary roughness is not merely the rugosity of the nonmoving bed and banks. Once the particles begin to move, the shear-resisting flow is altered. Particles can assume many different configurations, among which are dunes or ripples or a plane, and these

Propagation and harvesting. Strawberries are propagated by removing the runners, which form naturally, and setting them into new fields. *See* STEM (BOTANY). In many cases a new planting is made each year, produces fruit the following year and is then plowed under. However two annual crops may be produced from one planting, and sometimes as many as three or four. Harvesting is done by hand. A large part of the crop is sold fresh, a small amount made into preserves, and a large quantity frozen (*see* FOOD ENGINEERING).

Most varieties produce their crop in early summer, the picking season lasting about 2 weeks in the North, but much longer in the South In Florida, and to some extent in the other Gulf states, plants may be set in the fall and produce fruit throughout the winter (*see* REPRODUCTION, PLANT). In California the picking season extends over several weeks. Other varieties, usually limited to home garden production, produce flowers more or less continuously in the North, hence the term "everbearing" strawberry. *See* FLOWER (BOTANY).

Strawberry breeders have been active since the earliest introduction of the present cultivated fruit, all of our important varieties now being the result of controlled crosses. *See* BREEDING (PLANT). Strawberry varieties are usually adapted to rather limited areas so that different varieties are grown in each producing section, although a few seem adapted to more than one area. Varieties commonly grown in the South will produce during the short days of winter, whereas northern varieties will not fruit normally during short days (*see* PHOTOPERIODISM IN PLANTS). [J.H CE.]

Strawberry diseases. Strawberries are afflicted with root rots, leaf spots, and fruit rots, which vary in geographic distribution, destructiveness, and economic importance

Root diseases. Black root, most common of the fungus root rots, may be caused by species of *Fusarium*, *Rhizoctonia*, *Verticillium*, or by *Coniothyrium fuckellii*, and *Hainisia lythri* *See* FUNGI, ROOT (BOTANY). No single organism is implicated as the primary pathogen, but *Rhizoctonia* and *Fusarium* are the most frequent ones. *Verticillium* also causes a wilt of strawberry in England and in California

The most dangerous root disease, red stele, is caused by *Phytophthora fragariae*. Diseased plants have "rattail" root systems with few rootlets Root tips die and turn brown, while the central cylinder, or stele, of the root is red (*see* CORTEX, PLANT, STELE). A field once infested with the fungus is useless for strawberry production for many years Resistant varieties are available (*see* PLANT DISEASE CONTROL).

Nematodes cause root lesions and root knot and such diseases as spring and summer dwarf (*see* NEMATODA). Their importance in commercial production is not known.

Foliage diseases. Prevalent fungus diseases of strawberry are common leaf spot (*Mycosphaerella fragariae*), leaf scorch (*Diplocarpon earliana*), and leaf blight (*Dendrophoma obscurans*). Leaves, leaf and fruit stalks, stolons, calyxes, and fruit caps are attacked. Leaf blight occasionally causes fruit deterioration. *See* LEAF (BOTANY). Powdery mildew (*Sphaerotheca macularis*) is of minor importance in the United States. Leaf-spot diseases reduce yield and grade of marketable berries and weaken the runner plants.

Viruses occur in all commercial strawberry varieties and in wild species of Fragaria (*see* PLANT VIRUS). Certain virus strains and combinations of strains deform foliage and reduce plant vigor, stolon formation, and yield. Virus-free varieties can be grown.

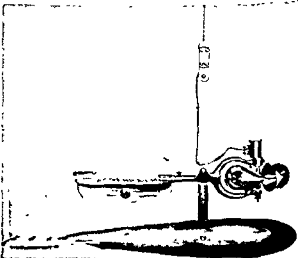
Fruit diseases. Greater losses result from fruit rots than from all other diseases combined Gray mold (*Botrytis cinerea*), tan rot (*Pezizella lythri*), leathery rot (*Phytophthora cactorum*), and brown rot (*Rhizoctonia* spp.) may be common and severe in the field. Leak (*Rhizopus nigricans*) probably is the most important market disease. [T.H K.]

Bibliography: *See* AGRICULTURAL SCIENCE (PLANT); PLANT DISEASE.

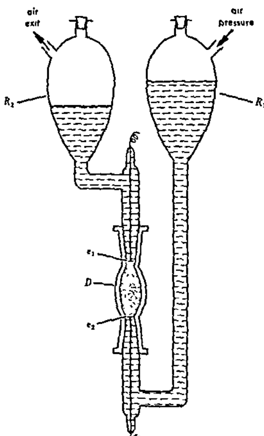
Stream gaging

The measurement of stream flow. A stream-gaging station is a particular site on a river where a record of stream flow is obtained. It usually consists of an instrument installation to record the fluctuating water level and also consists of facilities for making current-meter measurements of discharge. The discharge of a river is measured in the field by a direct observation of velocity and of the cross sectional area through which this velocity is applicable. By definition, discharge is the product of mean velocity times cross-sectional area. In English units, the velocity is expressed in feet per second, cross section in square feet, and thus discharge is cubic feet per second

Velocity is measured by a current meter, an instrument with vanes, blades, or rotor cups that turn a shaft at a rate depending on the speed of the water Each rotation of the shaft causes a click audi-



Price current meter and 30-lb C-type sounding weight. (U.S. Geological Survey)



Streaming potential apparatus

e_1 and e_2 is measured with an electrometer-potentiometer system This potential is the streaming potential E

In systems containing concentrations of electrolyte above 10^{-3} N, streaming potentials are too low to be measured accurately Then, the current produced by the streaming liquid may be used to evaluate the zeta-potential For capillaries of known dimensions, the following relationship for zeta-potential applies.

$$\zeta = \frac{4\pi\eta LI}{DAP} \qquad (2)$$

where I is the streaming current, L is the length of the capillary, and A is the cross-sectional area of the capillary For porous solids of unknown capillary dimensions the ratio L/A in Eq (2) may be evaluated by measuring the resistance R of the diaphragm impregnated with a liquid of known electrical conductance κ The relationship is

$$L/A = \kappa R$$

The zeta-potentials obtained from Eqs (1) and (2) are valid only when the flow of the liquid through the diaphragm is laminar and when the radius of curvature of the pores is greater than the thickness of the double layer *See* ELECTROKINETIC PHENOMENA. ELECTROOSMOSIS. [Q.V.W.]

Streamline flow

A condition of fluid flow characterized by the absence of turbulence. Other designations employed are laminar flow or viscous flow.

Fluid flow particles in streamline flow follow well-defined continuous paths or streamlines. At a fixed point in streamline flow, the flow velocity either remains constant (steady flow) or varies in a regular fashion with time (unsteady flow). In turbulent flow the velocity at a given point exhibits irregular, high-frequency fluctuations with time. In some instances, streamline flow can best be depicted as formed from thin layers of fluid which slip past each other (lamellar flow). As an illustration, streamline flow in a straight pipe might be considered as formed from layers in the shape of concentric annuli. If the flow can properly be represented by thin, plane layers (laminae) sliding past each other, the flow is commonly referred to as laminar flow although this term is often used to designate streamline flow in general.

The persistence of streamline flow in a given system is largely dependent on the value of a nondimensional parameter called the Reynolds number, defined as $\rho LU/\mu$ where ρ is the fluid density, U is a reference flow velocity, L is a reference length, and μ is the coefficient of viscosity. When the Reynolds number of a particular flow exceeds a certain value (the critical Reynolds number), transition from streamline flow to turbulent flow generally takes place

Under special circumstances flow may alternate between streamline flow and turbulent flow. This phenomenon is readily observed in pipe flow when the Reynolds number approaches the critical value. The pipe flow may become locally turbulent, the turbulent region passing downstream as a plug followed by streamline flow. As the plug leaves the pipe, the entire process is repeated

It is possible for both streamline flow and turbulent flow regimes to exist simultaneously in fluids with low values of viscosity, such as air. In such fluids, frictional effects are often confined to thin layers of fluid, known as boundary layers, in the immediate neighborhood of the bounding surfaces. The boundary layer may be locally turbulent while the flow in the mainstream away from the surfaces may be streamline flow. *See* BOUNDARY LAYER FLOW, FLUID MECHANICS; LAMINAR FLOW; STREAMLINING, TURBULENT FLOW. [A G HA.]

Bibliography: S Pai, *Viscous Flow Theory*, vol 1, 1956; L. Prandtl and O G. Tietjens, *Applied Hydro- and Aeromechanics*, 1934; H F. P. Purday, *Streamline Flow*, 1949.

Streamlining

The contouring of a body to reduce its resistance to motion through a fluid. The resistance to motion is referred to as the drag of the body.

Drag forces. Streamlining of a body must take into account the nature of drag forces Drag forces are of four types: induced drag, wave drag, pres-

bed forms depend on the transportation process. Thus the resisting shear at the boundary depends on the debris transport itself. *See* CHANNEL, OPEN.

When shear applied by water to a grain-bed of uniform-size particles becomes sufficient to move a layer of grains, successive layers do not progressively peel off indefinitely. After some layers are put in motion, an equilibrium is reached. Transport then continues without further degradation. R. A. Bagnold showed by theory and experiment that the grains in transport add a new force normal to the bed which holds the particles exposed at the bed against the stress of the overlying fluid-grain mixture. This force, the dispersive stress between sheared grains, makes a fundamental difference in the stress structure between fixed and movable-bed channels

Despite important advances in theory since 1950, it is possible to compute only approximately the rate of transport in open channels from hydraulic parameters and physical measurements of the channel and the debris Though the available methods give reasonable results in some instances, consistently acceptable results are not yet obtainable

From considerations of probability analysis and fluid mechanics H A. Einstein published in 1950 a formula and a computational technique to estimate total debris load of a stream The scheme presented certain difficulties in application owing in part, at least, to the fact that not all of the needed parameters are measured, to the use of certain empirical relations which are not adequately defined, to tedious computational procedures, and to the fact that the results obtained were often widely at variance with measured load when direct checks on the procedure were possible

In an attempt to make the Einstein equations more amenable to practical use, field engineers of the U.S. Geological Survey, using available direct measurements of total load, developed a simplified approach known as the modified Einstein proce-

ing problems involving sediment transport, even in the knowledge that the results are not completely satisfactory.

A significant advance in theory of sediment transport is the 1956 paper of R. A. Bagnold. His success in formulating a rational theory is indicated by the fact that the general equations for transport of unigranular (single size) cohesionless particles fit the transport of sand grains by wind and the transport of grains of a variety of densities in water, as well as some slurries But the Bagnold equations cannot be applied directly to computation of total load in natural streams because they treat only a grain-bed of uniform size.

Research is needed to learn more about the interaction effects of grain-mixtures and to adapt such knowledge into computational procedures. Also, the effects of cohesion and binding caused by silt and clay in stream debris must be studied. Research in stream transport has assumed new impetus in the decade beginning with 1950, and rapid advancement in the field may be expected. *See* FLUVIAL EROSION LANDFORMS; SEDIMENTATION (GEOLOGY)

[L.B.L]

Bibliography: R. A. Bagnold, The flow of cohesionless grains in fluids, *Phil. Trans Roy. Soc London, ser. A*, 249(964):235–297, 1956; B R. Colby and C. H. Hembree, *Computations of Total Sediment Discharge Niobrara River near Cody, Nebraska*, USGS Water Supply Paper 1357, 1955; H. A. Einstein, *The Bed-Load Function for Sediment Transportation in Open Channel Flows*, USDA Tech. Bull. 1026, 1950; G. K. Gilbert, *Transportation of Debris by Running Water*, USGS Profess. Paper 86, 1914.

Streaming potential

The potential which is produced when a liquid is forced to flow through a capillary or a porous solid. G. H Quincke (1859) found that the electromotive force produced by the streaming of pure water under a given pressure through a clay plate is independent of the size and thickness of the diaphragm and of the amount of water forced through the diaphragm; the electromotive force is, however, proportional to the pressure.

The streaming potential is one of four related electrokinetic phenomena which depend upon the presence of an electrical double layer at a solid liquid interface This electrical double layer is made up of ions of one charge type which are fixed to the surface of the solid and an equal number of mobile ions of the opposite charge which are distributed through the neighboring region of the liquid phase. In such a system the movement of liquid over the surface of the solid produces an electric current, because the flow of liquid causes a displacement of the mobile counter ions with respect to the fixed charges on the solid surface The applied potential necessary to reduce the net flow of electricity to zero is the streaming potential.

The principal objective of streaming potential measurements is the evaluation of zeta-potentials at solid-liquid interfaces. The relationship which may be used for this purpose is

$$\zeta = \frac{4\pi\eta\kappa E}{PD} \qquad (1)$$

where ζ is the zeta-potential E is the streaming potential, η is the viscosity of the liquid. κ is the conductance of the liquid as it exists in the capillary system. P is the applied pressure, and D is the dielectric constant of the liquid.

An apparatus used by R. A Gortner for measurement of streaming potentials at cellulose-water and alumina-organic liquid interfaces is shown in the illustration. Perforated gold or platinum electrodes, e_1 and e_2, are located on either side of a pad of compacted powder or fibers of a selected solid in diaphragm D. Liquid is forced by compressed air to flow from reservoir R_1 through the solid and into reservoir R_2. The potential between the electrodes

being joined by straight lines. Because the intensity of the shock wave and the drag is dependent upon the magnitude of change in direction, the nose and tail are as sharp as is practicable and the width or thickness of the body is minimal

When a body, streamlined for subsonic speed, operates at supersonic speeds, the blunt nose of the body causes the *impinging air to turn at too great an angle to sustain an attached oblique shock wave, as in Fig. 3* Thus, a strong shock wave, normal to the initial flow direction, is formed ahead of the body (Fig 4). The region between

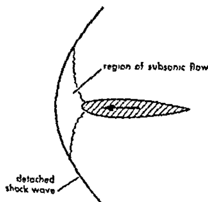


Fig 4. Flow about a body streamlined for subsonic flight traveling at supersonic speed.

this detached normal wave and the body is made up of air now moving at subsonic speeds and flowing around the blunt nose in characteristic subsonic fashion The intensity of the shock wave is much greater than the intensity of the waves shown in *Fig 3* The changes in pressure, density, and velocity are also large and the energy loss and corresponding wave drag are much greater than for a body designed for supersonic flight. *See* AIRPLANE, SHOCK WAVE, SONIC BARRIER; SUPERAERODYNAMICS [R.E BO.]

Bibliography· F J Bayley, *Introduction to Fluid Dynamics*, 1958; R A Dodge and M J. Thompson, *Fluid Mechanics*, 1937, S. Pai, *Introduction to the Theory of Compressible Flow*, 1959.

Strength of materials

A branch of applied mechanics concerned with the behavior of materials under *load, relationships between* externally applied loads and internal resisting *forces*, and associated deformations Knowledge of the properties of materials and analysis of the forces involved are fundamental to the investigation and design of structures and machine elements (see STRUCTURAL MATERIALS; MACHINE DESIGN) Mathematical application of principles of mechanics is supplemented by experimentally determined properties of materials and other empirical constants.

Investigation of the resistance of a member, dealing with internal forces, is called *free-body* analysis In it, principles of statics are applied to imaginary isolated segments of the loaded member (*see* STATICS). Determination of the distribution and intensity of the internal forces and the associated deformations is called stress analysis. *See* STRESS AND STRAIN.

Internal reactive forces developed *in response to* straining actions depend on the magnitude and nature of the loads. The four possible straining actions are (1) tension or compression, which lengthen or shorten the member; (2) shearing, which produces sliding or angular distortion along the plane of applied tangential forces; (3) bending, in which couples or bending moments produce change in curvature; and (4) torsion, in which couples acting normal to the axis twist the member. *See* BENDING MOMENT; SHEAR; TORSION.

A material offers resistance to external load only insofar as the component elements can furnish cohesive strength, resistance to compaction, and resistance to sliding. The relations developed in strength of materials analysis evaluate the tensile, compressive, and shear stresses that a material is called upon to resist. The most important factors in determining the suitability of a structural or machine element for a particular application are strength and stiffness.

Applications of fundamentals can be broadly classified as (1) investigation of members with known dimensions and materials to determine their ability to resist prescribed loads without excessive deformation, instability, or fracture, and (2) selection of suitable materials and determination of shape and dimensions of a member to perform a prescribed function involving known or estimated external loads. Design is the prediction of suitability for a prescribed function. [W J.KR.]

Streptococcus

A genus of bacteria of the family Lactobacillaceae. Microorganisms of the genus *Streptococcus* are well distributed in nature to include many strains which are pathogenic, or disease-producing, for man, other mammals, *birds*, and insects. There are strains which are generally harmless parasites but can cause severe infections under special circumstances which apply to the host. Some are consistently harmless, *and* a few are decidedly helpful The combination of ubiquity and range of activities explains the sustained attention these bacteria receive. *See* LACTOBACILLACEAE.

Morphology. Streptococci are seen und[illegible]

[illegible] plane with an inherent proneness of the units to remain attached. This typical chain formation and the round to ovoid shape of the single cell will quite accurately identify the *Streptococcus*. The bacteria are gram-positive in that they retain the blue color of Gram's differential bacterial stain

sure (or form) drag, and skin friction. Induced drag is usually important for finite lifting surfaces such as wings; wave drag is important for bodies moving at supersonic speeds. Bodies moving at subsonic speeds are influenced by pressure drag and skin friction.

Pressure drag is caused by inequalities in pressure forces acting in the direction of motion of the body. A body moving with uniform velocity through an ideal fluid (incompressible and nonviscous) experiences no pressure drag (*see* D'ALEMBERT'S PARADOX). In a real fluid, however, the action of viscosity tends to cause the flow to separate from the surface of the body with the consequent formation of a region of swirling or eddy flow termed the body wake. This eddy formation leads to a reduction in the downstream pressure on the body and hence gives rise to a force opposite to the body motion.

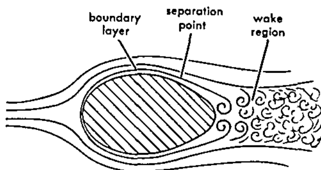


Fig. 1. Flow regions about a body in uniform subsonic flow.

Skin friction drag results from the frictional dissipation of energy due to fluid viscosity (*see* SKIN FRICTION). For fluids with relatively low viscosity such as water and air, effects of viscous friction are confined to a thin layer of fluid on the surface. *See* BOUNDARY-LAYER FLOW.

Under the influence of pressure gradients opposite to the direction of motion, the flow within the boundary layer tends to reverse and flow in an upstream direction. As a result, flow separation, mentioned earlier, occurs. Figure 1 shows the boundary-layer region, separation point, and wake region behind a body in uniform flow

Subsonic streamlining. In general, the streamlining of a body in subsonic flow is the contouring of the body in such a manner that the wake is reduced to a minimum. Drag is then mainly the result of skin friction. Because one of the main causes of flow separation, and hence wake formation, is rapid deceleration of flow along the body, the contouring must make deceleration gradual. These considerations lead to the following general rules for streamlining: (1) the forward portion of the body should be well rounded, and (2) the body should curve back gradually from the forward section to a tapering after-section with the avoidance of sharp corners along the body surface. These conditions are well illustrated by teardrop shapes (Fig 2). [A C.HA.]

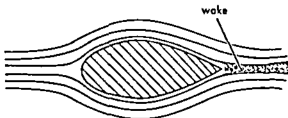


Fig. 2. Flow about streamlined body at subsonic speed.

Supersonic streamlining. At supersonic speeds the airflow can accommodate sudden changes in direction by being compressed or expanded Unlike subsonic flow, the air does not change direction until the most forward point of the body has passed. This is because the body is traveling faster than the speed at which pressure waves propagate in air (speed of sound). Where this change in direction first occurs, a compression wave is created, the strength of which depends upon the magnitude of change in direction, which in turn depends on the angle or sharpness of the nose of the body The sharper the nose, the less the change in direction and the weaker the compression shock wave

When the flow changes direction again at the midpoint of the body, the air will expand to follow the shape of the body (Fig 3). This change in direction creates an expansion shock wave At the tail of the body the direction changes again, creating a compression shock wave. At each of these shock waves, changes in pressure, density, and velocity occur and in this process energy is lost. This energy loss results in wave drag.

Bodies which are streamlined for supersonic speeds are characterized by a sharp pointed nose, a sharp pointed tail, and a minimum number of direction changes between the nose and the tail This requirement for a minimum number of changes in direction results in the nose and tail

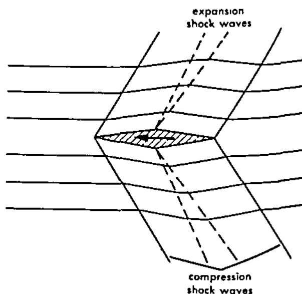


Fig 3 Shock waves about a streamlined body at supersonic speed.

or substrate and aerial mycelium. They occur abundantly in soil and in other natural substrates. They produce a variety of enzymes (proteolytic, diastatic, oxidative), vitamins (B_{12}), and antibiotics. Most of the important antibiotics isolated since the discovery of penicillin are produced by members of this family. It is sufficient to mention streptomycin, chloramphenicol, the tetracyclines, neomycin, erythromycin, novobiocin, nystatin, candicidin, and oleandomycin. *See* ANTIBIOTIC; ENZYME.

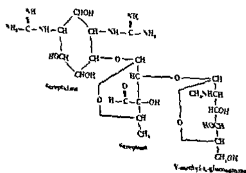
This family comprises at present several distinct genera, which are classified into two main groups—genera that produce aerial mycelium, and genera that, as a rule, do not. In the former group are (1) *Streptomyces*, which forms spores in chains, (2) *Thermoactinomyces*, which forms single spores, (3) *Waksmania* (*Microbispora*), which forms paired spores and grows best at 25–40°C (mesophile), and (4) *Thermopolyspora*, which grows best above 50°C (thermophile).

In the group that do not usually produce aerial mycelium are *Micromonospora*, which is a mesophile, and *Thermomonospora*, a thermophile. They form single spores on short sporophores. *See* ACTINOMYCETALES; BACTERIA, TAXONOMY OF.

[S A W]

Streptomycin

A colorless antibiotic substance produced by certain species of *Streptomyces*, mainly *S. griseus*. The antibiotic is produced in submerged culture in a medium consisting of protein-rich materials, such as soybean, cotton-seed meal, or peanut meal, together with some sugar or sugar-rich materials, such as distiller's solubles, as well as certain supplementary minerals. The length of the fermentation period is 2–3 days. At the completion of the fermentation period, the broth is filtered off [illegible] ion-exchange resins have supplanted [illegible] chloride complex salt from methanol solution. The calcium chloride complex salt and another form, streptomycin sulfate, in which streptomycin is commercially produced are amorphous white powders. In 1956, 622,000 lb of streptomycin were produced in the United States.



Streptomycin's chemical formula is $C_{21}H_{39}N_7O_{12}$, the molecule being composed of streptidine, streptose, and *N*-methyl-L-glucosamine, joined to one another by glucosidic linkages. Streptomycin is basic in nature, soluble in water, and thermostable. It is reduced chemically to dihydrostreptomycin, giving a preparation that is frequently used in clinical practice in preference to streptomycin itself.

Antimicrobial spectrum. Streptomycin and dihydrostreptomycin are active against gram-positive, gram-negative, and acid-fast bacteria, such as mycobacteria. These compounds have been used widely to treat infections of man, animals, and plants, such as tularemia, brucellosis, granuloma inguinale, fowl typhoid, plague, certain respiratory infections, and tuberculosis (*see* BRUCELLOSIS; GRANULOMA INGUINALE; PLAGUE; TULAREMIA). Streptomycin has also been effective in the following: bacteremia, meningitis, and urinary tract infections due to streptomycin-sensitive strains of various gram-negative bacteria; *Hemophilus influenzae* infections of the blood stream and heart, meninges, and respiratory and urinary tracts; endocarditis due to penicillin-resistant organisms, pneumonia due to *Klebsiella pneumoniae*; empyema and other pulmonary infections due to streptomycin-sensitive organisms; chancroid; granuloma inguinale; gonorrhea unresponsive to penicillin; glanders, corneal ulcers caused by *Pseudomonas aeruginosa*; wounds and cutaneous infections due to *Proteus vulgaris*; peritonitis and enteritis, and prophylactically in gastrointestinal surgery. *See* ENTEROBACTERIACEAE; KLEBSIELLA PNEUMONIAE; MENINGITIS; PSEUDOMONAS AERUGINOSA.

Streptomycin is not active against rickettsiae and viruses (*see* RICKETTSIOSES; VIRUS). Although as a rule it is not active upon fungi, it has found extensive application in the treatment of certain plant diseases caused by fungi, such as the blue mold of [illegible] and in the feeding of animals.

Organisms sensitive to streptomycin can become resistant to it upon continued contact either in culture media or in the animal body.

Use in tuberculosis. Streptomycin and dihydrostreptomycin are most effective in the treatment of various forms of tuberculosis (see TUBERCULOSIS). The discovery of streptomycin was announced early in 1944. Before the end of that year, W. H. Feldman and H. C. Hinshaw of the Mayo Clinic demonstrated its effectiveness in experimental tuberculosis of guinea pigs. Streptomycin was soon being submitted to clinical trials. The treatment of tuberculosis with streptomycin has been undergoing considerable change since its usefulness was first demonstrated. Its therapeutic value is limited by the fact that after exposure to streptomycin for

and are not decolorized during the staining procedure to absorb the pink color of the counter dye. In addition they are not dissolved by bile or bile salts. Thus, total appearance as to shape, arrangement, and color, with resistance to the action of bile by test, places microorganisms in the genus *Streptococcus*. *See* GRAM'S STAIN.

Oxygen relationship. Streptococci are usually aerobic and can be readily grown on culture mediums under ordinary atmospheric conditions. There are anaerobic strains which demand a diminished amount of oxygen for growth and some strains which are intermediate. The aerobic strains are the most numerous and the most important. Also more is known about them since study is not limited by the factor of growth.

Growth media. Streptococci multiply readily when the growth medium is adequate. Beef heart infusion agar with rabbit's blood and its liquid equivalent without agar will readily support the growth of streptococci. This is a practical medium but actually not a simple one in view of the contained elements. However, they are necessary because the requirements of streptococci are high when compared to microorganisms in general.

Growth on solid media. Discrete colonies grow on the surface of a blood agar medium after 18–24 hours of incubation at 37.5°C Each colony measures 1–2 millimeters in diameter and is grayish to greenish white. Streptococcal colonies are described as matt or glossy. Glossy varieties are more smooth and white, while matt colonies are rough and gray. After two or more days the edge of the matt colonies may lift and curl from the medium to give a cigarette-ash appearance.

Growth in liquid media. Multiplication of organisms in beef heart infusion broth is either diffuse (cloudy) or granular Granularity of appearance results from the formation of very long chains which become entangled. The rabbit blood cells at the bottom of the tube may be unchanged, appear purple, or be quite completely destroyed to give the broth a wine color. This change in appearance of the blood is also reflected in the halo surrounding a colony on the agar plate.

Classification. Streptococci are classified by their influence on the red blood cells of the growth medium, fermentation of sugars and other biochemical reactions, and by the serologic or immunochemical method. No system is complete, although serologic classification approaches this goal. The three methods partially overlap and have a varying practicability. It is common to employ one or two methods to the degree needed by the problem.

Important species. *S. hemolyticus* completely dissolves red blood cells and is considered by many as the most important of the streptococci. Septic sore throat, peritonsillar abcess (quinsy), scarlet fever, erysipelas (St. Anthony's fire), puerperal sepsis (childbed fever), discharging ear, and visibly swollen neck glands are a few of the clinical pictures produced by those hemolytic streptococci of Lancefield group A, which can be further classified into about fifty types.

S. hemolyticus produces a number of substances during cultivation. These have a deleterious action on the blood and other body tissues of the host. It is felt that as a group they aid the microorganisms both to infect and to spread from the local lesion. One such substance, which has been studied in much detail, is the erythrogenic, or scarlet fever, toxin. It is produced by all strains in varying amounts, but its effect on the human being is influenced by previous experience with the bacterium. Scarlet fever toxin has been used in standard amounts to immunize against this one clinical variety of streptococcal disease and to determine existing immunity in the Dick test. In this test an absence of skin redness at the site of injection indicates that scarlet fever will probably not result. Such a negative Dick test occurs in two situations. Older children and adults usually give no reaction since they have scarlatinal antitoxin or antibody in their blood from previous scarlet fever or from repeated streptococcal infections without the rash. Newborns and many older infants likewise fail to react, and scarlet fever is rare at this early age. This resistance is on another basis since the infant does not possess antitoxin and the result is independent of the mother's response to the Dick test. It has been suggested that a hypersensitivity must be developed by prior exposure to the erythrogenic toxin before the characteristic rash can be induced. Thus, the toxic action depends on two different factors, that is, the state of hypersensitivity and the absence of antitoxin. In addition, some streptococci produce a skin toxin which is different from that produced by almost all the others. Second attacks of scarlet fever can be explained by this. *See* ANTIBODY.

Epidemiology. The difference in contagiousness of streptococci is not understood. Infection of babies and the presence of organisms in the nose of any carrier, regardless of age, are recognized as important factors.

Streptococci do not develop resistance to the antibiotic drugs such as penicillin. Resistance by disease-producing bacteria is deemed an added means of ensuring continued existence of the particular bacterial race. Streptococci appear to rely on the ability to survive and spread from person to person with the production of little or no demonstrable illness. *See* BLOOD PLATE HEMOLYSIS; EPIDEMIOLOGY; LANCEFIELD DIFFERENTIATION SCHEME; MASTITIS (COWS); RHEUMATIC FEVER; SCARLET FEVER; SKIN TEST. [P.L.B.]

Bibliography: R. J Dubos (ed.), *Bacterial and Mycotic Infections of Man*, 1958.

Streptomycetaceae

A family of bacteria of the order Actinomycetales. The Streptomycetaceae comprises aerobic actinomycetes that usually produce a typical vegetative

cise, or exposure to anoxia, to extremes of temperature, or to chemical or surgical insult, responses of organisms to such stresses can be compared in terms of quantitative measures of their responding regulatory systems. Thus, following exercise or a fight, elevation and subsequent recovery to normal levels of respiratory rate, heart rate, blood pressure, body temperature, blood acidity, oxygen consumption and carbon dioxide (CO_2) production may be used as measures of strain resulting from the stress.

Blood and urinary measurements of certain endocrine systems brought into play by stress are particularly useful response indices. The quantitative analyses of adrenalin, secreted primarily by the medulla of the adrenal gland, and noradrenalin, a neurohumor primarily released at certain synapses and nerve endings of the autonomic nervous system, reflect defense responses especially to acute stress (*see* ADRENAL GLAND). Associated with some of these stresses may be expressions of rage or fear. The ratios of adrenalin and noradrenalin vary with the nature of the stress and with the type of response elicited. In man, the excretion of adrenalin in the urine is primarily enhanced by situations involving tense, anxious, nonaggressive emotional responses, while noradrenalin tends to be excreted more in situations calling for actively aggressive, combative responses.

Blood and urine measurements of the steroid hormones from the adrenal cortex and their metabolites, such as the 17-ketosteroids, together with certain blood and urinary constituents reflecting actions of these hormones on target organs in the body to produce eosinopenia, lymphopenia, ...d changes ...

d, ha...

lices ...

...onses in mammals. The adrenal cortex is acti...ted by the pituitary adrenocorticotrophic hor...one (ACTH) which, in turn, is released in in...eased amounts by action of the hypothalamus, ...llowing bodily damage or threats. The adreno...rtical hormones have ubiquitous actions on many ...ssues involved in response to stress and the main...nance of homeostasis. There is some tendency to ...e adrenocortical responses essentially as meas...res of stressful situations that cannot otherwise be ...uantitated. This, however, has led to confusion. ...ince there are differences between individuals in ...drenal response, not only in terms of the patterns ...f specific secreted adrenal corticoids and their metabolites but also in the correlations of adreno...cortical indices with other physiological and psy...chological responses to stress. *See* HORMONE, STEROID.

Systematization of terminology. Hans Selye, in 1950 systematized the language relating to stress in the biological sciences. He refers to an adverse influence that acts upon an organism to produce a condition of stress as a stressor or stressor agent. Thus surgical trauma, hemorrhage, infection, the ingestion of toxic agents, anoxia, intense or prolonged exercise, exposures to extremes of heat and cold, and exposure to psychologically disturbing situations are stressors. The reaction of the organism to such stimuli is the stress response and examples of such responses have been given above. Stress is defined by Selye as the state of an organism subjected to a sufficiently powerful stressor so as to result in damage to the organism or to call forth its defense reactions.

Selye has further developed the concept of the general adaptation syndrome. He defines this as the characteristic emergency reaction or general stress response of an animal that develops through three stages: (1) the alarm reaction in which adaptation is attempted, (2) the stage of resistance in which adaptation is optimal, and (3) the stage of exhaustion in which adaptation fails. These various phases of the general adaptation syndrome may be studied in terms of the sort of changes in constituents of body fluids discussed above, especially those involving activity of the adrenal cortex. In animals, the studies may include determinations, before and after stress, of adrenal size, adrenal ...

...the bodily mechanisms for adaptation adequately to meet chronic stress situations. He considers that disturbed patterns of endocrine secretion following the prolonged application of stressors may result in various chronic diseases and he has thus spoken of the diseases of adaptation resulting from prolonged stress.

Prolonged stressor actions may produce anxiety states and chronic behavioral disturbances in man and in experimental animals. The role of life stresses as contributory agents to hypertension, arthritis, ulcers, skin disorders, asthma, and other allergic manifestations has been investigated by many biochemical, physiological, psychological, and psychiatric procedures. Experimentally induced states in animals, closely resembling the neuroses and psychosomatic disturbances seen in man, have been regularly brought about by frustrating conditioned reflex techniques, in themselves mild procedures but productive of highly abnormal and crippling behavior when carried on at regular intervals over periods of time. Life stressors producing neurotic behavior in man are very varied and may be difficult to define or to measure. In the biosciences, stresses for the most part are not directly identifiable but are measured by the strains they produce, some of which may exceed the elastic limit or adaptive ability of the organism and so produce damage.

In relation to prolonged stress it has been suggested that nerve impulses, generated by maintained uncertainty in anxiety-inducing situations, may lead to reverberating and continuous passage of "messages" around looped systems. If such circuits are active for long periods, they may spread extensively in the nervous system and become permanent, even to the point of outlasting the induc-

weeks or months, strains of *Mycobacterium tuberculosis* resistant to the effects of the drug may be isolated. Therefore, streptomycin is used only as an adjunct to other measures in the treatment of tuberculosis, and is primarily of value in conditions in which temporary suppression of the infection will enable the patient to gain ascendency over the disease. Streptomycin should not be used in minimal and primary pulmonary tuberculosis, which will respond readily to routine treatment, since there is the danger that resistant organisms will emerge. Unnecessary use of the drug may interfere with its effectiveness when there is a more serious need.

In pulmonary tuberculosis, streptomycin is used mainly in combination with *p*-aminosalicylic acid (PAS) and isoniazid (INH) (*see* PARA-AMINOSALICYLIC ACID; ISONICOTINIC ACID HYDRAZIDE). The usual dosage is 1 gram (g) streptomycin a day +12 g PAS daily; or 1 g of streptomycin twice a week +12 g PAS daily; or 1 g streptomycin daily +300 mg INH daily. It is now common practice to use a mixture of streptomycin and dihydrostreptomycin in place of streptomycin alone. In this manner, the toxic effect of each substance is minimized.

Pharmacology. Streptomycin passes readily into the blood stream, following parenteral administration. Because of great individual differences in absorption and excretion of the drug, significantly different blood levels are to be expected in individual patients receiving the same dosage. The concentration of the drug in the blood varies directly with the size of the individual dose and frequency of administration. During the 24-hour period following intramuscular or intravenous injection, 60–80% of streptomycin is excreted by the kidneys, significant urinary levels may be obtained with comparatively small parenteral doses; these levels vary inversely with urinary output. Streptomycin passes readily from the blood into the peritoneal cavity. Following intramuscular injection, it diffuses also into the amniotic and intraocular fluids, and into the placental blood. In the absence of severe hepatic damage, it is excreted in the bile. If the cystic or hepatic duct is obstructed, no streptomycin will appear in the gall bladder. It does not diffuse readily into the pleura in most cases, nor have effective concentrations been demonstrated in the prostatic fluid following parenteral injection. Oral administration produces no detectable concentration in blood, a negligible amount in urine, and a high concentration in feces. [S.A.W.]

Stress (psychological)

A state or condition of an organism subjected to environmental forces which tend to upset steady-state and equilibrium conditions in the organism. The organism's reaction is to attempt to restore such conditions. The phrase "stresses of living" refers to the impacts on persons of unpleasant environmental and social events encountered in the course of daily life. The word stress, together with such terms as force and power, has precise quantitative meaning in physics but originally came from popular usage.

Stress in physics and engineering. In physics and engineering a stress is defined as a force applied to a solid body. It is measured as pressure per unit area causing deformation of the body. The deformation is defined as a strain and is expressed as a ratio of the change of length produced by the force to the original length measured along any axis. The ratio of stress to strain is known as the modulus of elasticity and is a constant characteristic of the body.

Stress in biology. In the biosciences a stress has been considered by many physiologists as any set of events which modifies steady state conditions within the organism so as to activate adaptive, or homeostatic, mechanisms. These mechanisms readjust the internal environment with the resultant reestablishment of normal steady state or equilibrium conditions that were modified by the stress. Thus, a shift in blood acidity following exercise calls upon both chemical buffering systems in the blood and neurophysiological mechanisms to reestablish homeostasis (*see* HOMEOSTASIS). External temperature changes activate autonomic, or involuntary, regulatory mechanisms to maintain constancy of the temperature of the internal environment, by eliminating body heat at a greater rate in a hot environment and conserving it in a cold one. The ingestion of chemical agents, foodstuffs in excess, or toxic compounds likewise activates homeostatic processes of elimination and inactivation. All of these events may be considered as stressful as they are reflected in measures of adaptive adjustment.

Any external situation threatening the organism may function as a stress. Situations calling for flight or fight, with their concomitant psychological and physiological expressions of fear and anger, are stressful. The processes of inhibiting fight or flight may themselves result in stressful anxiety states. Psychological stress may result from the intensification of instinctual drives and of the control of these drives to meet the demands of society. Such stresses may be chronic and produce far-reaching disturbances of a psychosomatic nature as well as neuroses in susceptible individuals (*see* NEUROSIS). The balancing of one's needs and satisfaction in terms of learned inhibitions and prohibitions represents stresses of this sort. The same stress situation may have quite different significance, psychologically, for different organisms in terms of their life histories and past conditionings and therefore attempts to objectify and standardize such stresses, per se, meet with great difficulty.

Measurement of stress. The physiologist has a variety of measures of how stress may disturb the body's regulation of the internal fluid environment of its tissues. The maintenance of constancy of this internal environment of blood and lymph is of great importance for proper functioning of the body. If, for example, the stress is a bout of exer-

change in volume. In axial tension, deformation is expressed by per cent elongation or per cent reduction of area to fracture and taken as a measure of ductility.

Stress-strain characteristics. Several terms are used to describe the strain behavior of materials in the presence of stress. Figures 1 and 2 describe graphically some of these characteristics.

Yield strength is the stress accompanying a specified permanent plastic strain, which is considered as not having impaired useful elastic behavior and which represents the practical elastic strength for materials having a gradual knee in the stress-strain curve. The offset or total extension methods utilize the stress-strain curve to evaluate the stress at which a specified plastic strain (usually 0.2%) or a specified total strain under load (as 0.5%) has developed *See* SAFETY FACTOR.

Proportional limit is the greatest stress a material can sustain without departure from linear proportionality of stress [illegible]

[illegible] materials such as low-carbon steel *exhibit* the unique phenomenon of sudden yielding The stress at which yield initiates is called the upper yield point The lower yield point is the constant stress while yielding progresses and is taken as a characteristic property of the material.

Ultimate strength defines the maximum resistance to tensile, compressive, or shearing forces, expressed either as a total load producing fracture as in the case of a *rope or cable*, the maximum stress developed prior to fracture, or the stress accompanying some limiting [illegible] strengt[illegible] at max[illegible] the breaking stress *See* BRITTLENESS.

Ultimate tensile strength is the maximum nominal tensile stress developed during increasing load application, calculated from maximum applied load and original unstrained sectional area Materials developing large elongation reach maximum load resistance prior to fracture, which occurs at a locally reduced section *See* STRESS CONCENTRATION

Ultimate compressive strength has meaning only when maximum load produces fracture as in brittle materials Metals that develop large deformation increase in sectional area, thus increasing resistance to load, they do not fracture

Modulus of rupture is a measure of strength in bending or torsion expressed as a maximum tensile compressive or shear stress computed from the maximum load to fracture and the dimensions of the member Its evaluation incorrectly assumes elastic behavior to fracture and therefore is not the true strength of the material It serves as an empirical measure of rupture strength useful in comparing quality of materials such as wood, concrete, and cast iron subjected to standard tests

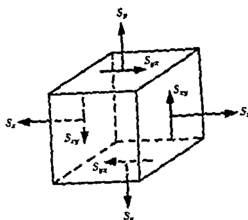


Fig 3. Combined stresses.

Combined stresses. Simultaneous action of stresses produced by independent straining actions such as tension and torsion, or bending and thrust produce combined stresses. The state of stress is defined by the magnitude and direction of normal and shearing stresses acting on an element of a structural member, the element having known orientation with the axes of the member. Stresses are three-dimensional when normal and shearing stresses act simultaneously on each face of a cubical element, and are two-dimensional or coplanar when stresses act in mutually perpendicular directions in the same plane. Stresses at a boundary free from surface stress are coplanar

The coplanar stress system shown in Fig. 3 has normal stresses S_x and S_y, and complementary shear stresses of equal intensity forming equilibrating couples. Stresses on planes of reference are found by conventional formulae for independent straining actions. Stresses on other planes are found by statics applied to a free body isolated from the element subjected to combined stresses. Planes on which maximum and minimum normal stresses occur are called principal planes, and these wholly normal stresses are principal stresses No shearing stresses exist on principal planes. Maximum shearing stress occurs on planes at 45° to the principal planes An element subjected only to principal stresses is in a state of biaxial stress; when one of these normal stresses is zero, the condition is uniaxial stress.

Expressed in terms of known normal stresses S_x and S_y and a shear stress S_{xy}, the principal stresses are

$$S_{\substack{max \\ min}} = \tfrac{1}{2}(S_x + S_y) \pm \sqrt{[\tfrac{1}{2}(S_x + S_y)]^2 + S_{xy}^2}$$

The planes on which principal stresses act are such that

$$\tan 2\theta = \frac{S_{xy}}{\tfrac{1}{2}(S_x - S_y)}$$

where θ is the angle with the reference axis of the element.

ing situation. If the spread of reverberating activity comes to include circuits in the hypothalamus, this may produce excessive stimulation of endocrine systems, especially that of the pituitary-adrenal axis. In cases of prolonged stress, the nature of the adrenal corticoid output itself may be changed, by shifting ratios of specific steroid hormone production. Excessive production of some steroids and shifts in hormone metabolism have chronic feedback effects on various organ systems, producing manifestations reflected in psychosomatic disorders and disorders of behavior. *See* ABNORMAL BEHAVIOR; EMOTION. [H.HO.]

Bibliography: H. Selye, *The Physiology and Pathology of Exposure to Stress*, 1950; *Fifth Annual Report on Stress*, 1952; Stress and mental illness, *Lancet*, 275:205–208, 1958.

Stress and strain

Related terms used to define the intensity of internal reactive forces in a deformed body and associated unit changes of dimension, shape, or volume caused by externally applied forces.

Stress is a measure of the internal reaction between elementary particles of a material in resisting separation, compacting, or sliding that tend to be induced by external forces Total internal resisting forces are resultants of continuously distributed normal and tangential forces of varying magnitude and direction and acting on elementary areas throughout the material. These forces may be distributed uniformly or nonuniformly.

Stresses are identified as tensile, compressive, or shearing, according to the straining action

Strain is a measure of deformation such as (1) linear strain, the change of length per unit of linear dimensions; (2) shear strain, the angular skew in radians of an element undergoing change of shape by tangential forces; or (3) volumetric strain, the change of volume per unit of volume. The strains associated with stress are characteristic of the material.

Strains completely recoverable on removal of stress are called elastic strains. Above a critical stress, both elastic and plastic strains exist, and that part remaining after unloading represents plastic deformation called inelastic strain. Inelastic strain reflects internal changes in the crystalline structure of the metal. Increase of resistance to continued plastic deformation due to more favorable rearrangement of the atomic structure is called strain-hardening

Stress-strain diagram. A graphical representation of simultaneous values of a stress and strain observed in tests indicates material properties associated with elastic and inelastic behavior. The diagram indicates significant values of stress accompanying changes produced in the internal structure.

Properties of metals are usually determined by tension or torsion tests. Materials such as wood, concrete, and ceramics that are used to resist compressive loads are tested under compression.

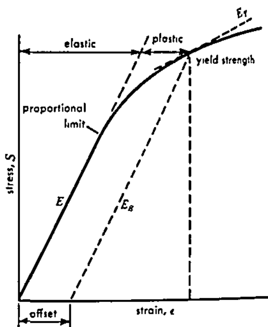


Fig. 1. Stress-strain diagram for an aluminum alloy

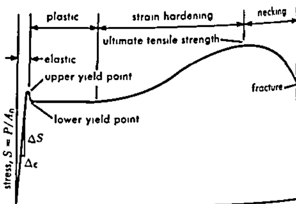


Fig 2. Stress-strain diagram for a soft steel

Tension tests are made on suitably machined bars in a testing machine which elongates the specimen and records the resisting force. The stress is the force per unit of original unstrained area of the cross section. Strain is the average unit elongation in the direction of the applied force, determined from total elongation of a selected gage length.

A tension stress-strain diagram for an aluminum alloy is shown in Fig 1. The diagram for soft steel, which exhibits the unique phenomenon of yielding at a critical stress, is shown in Fig. 2. Diagrams with stresses based on the original sectional area are called engineering or nominal stress-strain curves. A true stress-strain curve is based on the reduced sectional area accompanying an applied force The interpretation of these diagrams establishes characteristic properties of the material.

Deformation is any alteration of shape or dimensions of a body caused by stresses, thermal expansion or contraction, chemical or metallurgical transformations, or shrinkage and expansions due to moisture change. Deformation is measured by changes in linear dimensions angular skew, or

of the moving mass is transferred to the resisting body in the form of strain energy. The stresses produced by impact will depend upon the amount of energy transferred, type of straining action produced and the characteristics of the material.

The strain energy per unit volume stored in the material when fractured is represented by the total area under the stress-strain curve and is called the toughness. The strain energy stored while the strains are wholly elastic is called resilience. Maximum elastic strain energy per unit volume is called proof resilience and is a property of the material depending upon the elastic limit and the modulus of elasticity. *See* ELASTIC LIMIT; ELASTICITY.

The capacity to absorb energy under impact load is also measured by special tests in which the energy of a swinging pendulum or falling weight producing fracture is measured. *See* STRENGTH OF MATERIALS. [W.J.KR.]

Bibliography: H. E. Davis, G. E. Troxell and C. T. Wiskocil, *Testing and Inspection of Engineering Materials*, 2d ed., 1955.

Stress concentration

A condition in which a stress distribution has high localized stresses. A stress concentration is usually induced by an abrupt change in shape of a member. In the vicinity of notches, holes, changes in diameter of a shaft, or application points of concentrated loads, maximum stress is several times greater than where there is no geometrical discontinuity

sipate

the d

the se

The tensile stress distribution in a plate reduced by circular notches is shown qualitatively in the drawing. The stress at the root of the notch is about three times the stress at the

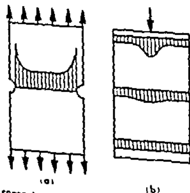
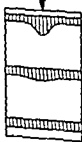
concentrate

formly distr

tions with t

sections, as

Stress concentrations are usually determined experimentally by photoelastic methods (*see* PHOTOELASTICITY). The peak stress is expressed as a multiple of the average or nominal stress computed without regard to the concentration, thus $S_{max} = KS_{nom}$, where K is the stress concentration factor, which depends only on the geometry, without regard to load magnitude. Factors have been evaluated for common types of discontinuity.

Stress concentrations. (a) Relieved plate under uniform tension. (b) Bar under concentrated end load

Where loads are steady and the material ductile, stress concentrations are relieved by plastic yielding and are not important. Under cyclic loads, fatigue cracks are initiated at points of high local stress, where most failures in machines occur. *See* STRESS AND STRAIN. [W.J.KR.]

Bibliography: M. M. Frocht, *Strength of Materials*, 1951; R. E. Peterson, *Stress Concentration Design Factors*, 1953; R. J. Roark, *Formulas for Stress and Strain*, 3d ed., 1954.

Striation

A succession of alternately luminous and dark regions sometimes observed in the positive columns of glow discharges. Usually striations appear when the discharge is operated at relatively high pressure, because diffusion effects will be less prevalent under this condition. The striations may be either stationary or moving, although these are thought to be separate phenomena.

The stationary striations are thought to be repetitions of the alternate luminous and dark spaces found in the cathode region. It is significant that very pure inert gases do not exhibit this effect. Their failure to do so is believed to be due to the presence of excited atoms in the metastable states characteristic of the inert gases. These metastable atoms can diffuse freely and can produce excitation uniformly throughout the column. The resulting decays will then result in uniform light production. A small amount of impurity gas can release the metastable states very quickly by collisions of the second kind. Thus, a small amount of impurity will permit the formation of striations. The moving striations are thought to be caused by plasma oscillations, and are not well understood. *See* GLOW DISCHARGE. [G.H.MI.]

Strigiformes

The order of birds which contains the

anatomical differences between the two groups. Both Tytonidae and Strigidae are widely distributed; the common barn owl (*Tyto alba*) is one of the most widespread bird species in the world. Each family includes both diurnal and nocturnal hunters; some of the latter have developed amazing adaptations for locating prey in total darkness, employing only hearing. Because of their predominantly nocturnal habits, many owls, particularly the tropical species, are little known other than as museum specimens. Similarities to the hawks, near which owls were formerly classified, are now attributed to convergence, both

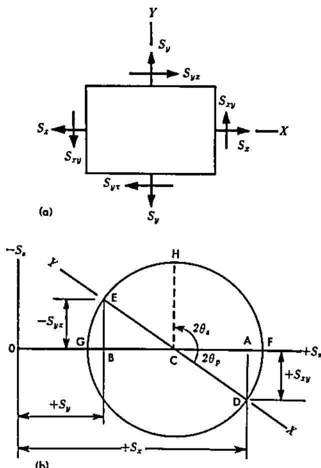


Fig 4. Mohr's construction. (a) Combined stresses. (b) Graphical construction.

Mohr's circle. A graphical construction called Mohr's circle determines the simultaneous combinations of coplanar normal and shearing stresses on any plane perpendicular to the plane of stress through a given point in a stressed body. The construction solves combined stress problems. For the state of stress shown in Fig. 4a, stresses on other planes are found by the construction in Fig. 4b. Referring normal and shear stresses to $+S_n$, S_s coordinate axes, OA = S_x and OB = S_y. Verticals AD and BE represent the equal shear stresses on faces having normal stresses S_x and S_y. Line ED is the diameter of Mohr's stress circle with center at C. Coordinates of a point on the stress circle represent corresponding normal and shearing stresses on a particular plane. Distances OG and OF are principal stresses. The angle which the maximum principal stress makes with the X reference axis is one-half angle DCF. Ordinate CH is the maximum shear stress acting on a plane defined by angle DCH = $2\theta_S$. Stress components on any plane whose normal makes an angle θ with the X axis can be found from the stress circle.

Bulk modulus is a constant designated K or E_v, it is a ratio of stress to accompanying volumetric strain, when the material behaves elastically. The volumetric strain ϵ_v is the change of volume per unit volume, $\epsilon_v = \Delta V/V$. According to Hooke's law, $K = E_v = S/\epsilon_v$ under stress S. For mutually perpendicular triaxial stresses, the volumetric strain is equal to the sum of the three linear strains and for hydrostatic stress $\epsilon_v = 3\epsilon$. For a body subjected to hydrostatic pressure

$$\epsilon_v = \frac{S}{E_v} + 3\left[\frac{S}{E}(1 - 2\mu)\right]$$

where

$$E_v = \frac{E}{3(1 - 2\mu)}$$

in which E is Young's modulus, and μ is Poisson's ratio.

Poisson's ratio μ is a material constant expressing the ratio of lateral strain to longitudinal strain (Fig. 5). For uniaxial stress

$$\mu = \frac{\text{lateral strain}}{\text{longitudinal strain}}$$

The ratio varies for different materials, usually ranging from 0.25 to 0.35. A maximum value of 0.5 represents plastic behavior.

The generalized Hooke's law for a body subjected to mutually perpendicular normal stresses including the Poisson's ratio effect is, for strain in the x direction,

$$\epsilon_x = \frac{s_x}{E} - \mu\frac{s_y}{E} - \mu\frac{S_z}{E}$$

Impact strength. Resistance of a material to dynamically applied loads, expressed as the capacity to absorb energy, in units of inch-pounds per cubic inch, is impact strength. An important dynamic load is one which is applied suddenly, as by the impact of a moving mass. The kinetic energy

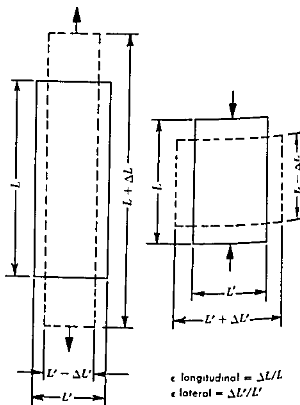


Fig 5. Longitudinal and lateral strains.

ıade to appear stationary. The object under view is een only when it is in one position in space and is reed of the blur normally associated with motion f the viewing rate is slightly greater than the repetition rate of the motion being observed, the object :ill appear to move slowly in reverse direction. imilarly, a slightly slower viewing rate makes the bject appear to move slowly in the direction of ıctual motion.

The flashing-light stroboscopes employ gas dis- 'iarge tubes to provide a brilliant light source of ry short duration A brilliant source is required ecause of the short period of illumination. Background illumination should be kept as low as possi- le Tubes may vary from neon glow lamps, when ery little light output is required, to special strob- scope tubes capable of producing flashes of several hundred thousand candlepower with a duration of only a few millionths of a second

Tubes may be fired by voltage peaking transformers or by electronic control circuits When eaking transformers are employed, the flashes are synchronized with the supply line Electronic controls provide adjustable or synchronous flashing frequencies Calibrating circuits give direct indication of the frequency of the flashes [A.R.E.]

Stroboscopic photography

Stroboscopic or "strobe" photography is generally understood to refer to pictures of both single and multiple exposure taken by flashes of light from electrical discharges. Originally the term referred to multiple-exposed photographs made with a stroboscopic disk as a shutter. One essential feature of modern stroboscopic photography is a short posure time, usually much shorter than can be tained by a mechanical shutter

High speed photography with stroboscopic light is proved to be one of the most powerful research ols for observing fast motions in engineering and science. Likewise, the electrical system of producing flashes of light in xenon filled flash lamps of great utility for studio, candid, and press hotography Spark photography, especially with hort air gaps at high voltage, has been used for many years to take short exposure photographs.

Energy storage. Devices for storing energy are required to produce the high peak power needed for producing pictures with a short exposure time The electrical capacitor is ideal for this service. The energy stored in a capacitor is

$$\frac{CE^2}{2} \quad \text{watt-seconds (or joules)}$$

where C is the capacitance in farads and E is the initial voltage to which the capacitor is charged. The electrolytic type of capacitor has the largest ratio of stored watt seconds per pound, but is limited to about 500 volts Paper capacitors are in widespread use for short-duration flash lights where high voltage is required or where the flash cycle is repeated frequently.

Flash lamps. A gaseous discharge lamp or spark gap is required to convert the stored energy of the capacitor into light. The open spark in air is used where a small volume source of very short duration is required Otherwise the more efficient xenon-filled flash lamp is in almost universal application since its efficiency is about five times greater than that of the open spark. There is an afterglow in xenon gas of about 10 μsec or more which is objectionable when fast subjects such as high-speed bullets are photographed.

Flash duration. Light duration from a xenon lamp or a spark gap is influenced by the electrical characteristics of the capacitor, the series inductance of the capacitor and leads, and the resistance of the leads. For air gaps, the effective resistance of the gap is negligible compared to these other factors. However, a long xenon-filled flash lamp of small diameter has a high resistance that limits the current flow The flash duration of a xenon lamp can be much longer than for an air spark.

Circuit of a flash lamp. Figure 1 shows an electronic flash lamp L such as a xenon-filled glass tube with one external and two internal electrodes. The lamp is connected directly across the main flash capacitor C which stores the energy that is to be converted by the lamp into light. Current from the transformer, after being rectified, charges the capacitor C to a voltage equal to the peak voltage of the secondary of the transformer T.

An electrical discharge is started in the flash lamp when switch S is closed. An X contact in a synchronized camera shutter can serve as the switch S The pulse voltage in the secondary of the step-up transformer T_2 is connected to the external electrode on the lamp and starts a current in the tube by condenser action. Once the glow on the

[illegible]

... and then slowly decays to zero. The transient is a complex one to analyze because the gaseous discharge is nonlinear. See ELECTRICAL CONDUCTION IN GASES

The complete history of a discharge, showing light as well as current and voltage against time, must be measured experimentally for the exact

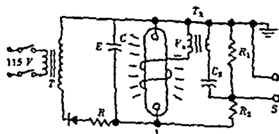


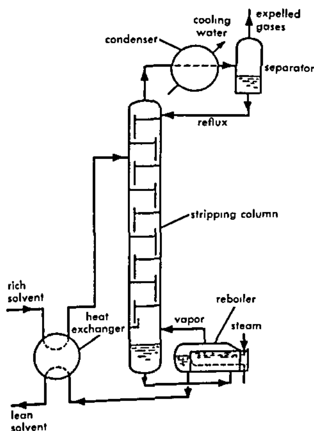
Fig. 1. Elementary circuit of an electronic flash lamp L, with main flash capacitor C and trigger circuit with transformer T_2. The lamp L flashes when the synchronizing switch S is closed

groups being wholly carnivorous or insectivorous. All owls lay the white eggs typical of hole-nesting birds, but a few species nest on the ground or in old hawk or crow nests. Incubation is usually by the female which is the larger member of the pair. *See* AVES. [K.C.P.]

Stripping

The removal of a volatile component from a liquid by vaporization. The stripping operation is an important step in many industrial processes which employ absorption to purify gases and to recover valuable components from the vapor phase (*see* GAS ABSORPTION OPERATIONS). In such processes, the rich solution from the absorption step must be stripped in order to permit recovery of the absorbed solute and recycle of the solvent.

The stripping of a volatile component from a liquid may be accomplished by pressure reduction, the application of heat, or the use of an inert gas (stripping vapor). Many processes employ a combination of all three; that is, after absorption at elevated pressure, the solvent is flashed to atmospheric pressure, heated, and admitted into a stripping column which is provided with a bottom heater (reboiler). Solvent vapor generated in the reboiler or inert gas injected at the bottom of the column serves as stripping vapor which rises countercurrent to the downflowing solvent When steam is used as stripping vapor for a system which is not miscible with water, the process is called steam stripping. *See* DISTILLATION.



Schematic diagram of stripping column and accessory equipment.

Equipment used for stripping operations resembles that employed for absorption and distillation and consists generally of countercurrent columns of the bubble-plate or packed types The rich solvent is admitted near the top so that the major portion of the column is active for stripping; however, in many cases, a short section is provided above the feed point and condensate is refluxed to this section to minimize losses of solvent or to return volatile components which are to be retained in the liquid phase. Stripping columns are also referred to as strippers, desorbers, regenerators, reactnators, and stills (if heat is used).

The operation of stripping is essentially the reverse of absorption in that mass transfer occurs from the liquid to the gas phase. This requires that the partial pressure of solute in the gas be less than the equilibrium vapor pressure of solute over the solution (or, with reference to the liquid, the concentration of solute in the liquid must be greater than the concentration in equilibrium with the gas phase). The stripping operation is also closely related to distillation and in some instances, such as the stripping of hydrocarbons from an absorption oil, the process can be considered the equivalent of a distillation operation in which the bottom and top products have widely different boiling points

It should be noted that the term stripping has other technical meanings besides that discussed above, including the removal of organic or metal coatings from solid surfaces and the removal of color from dyed fabrics. *See* DYEING; ELECTROPLATING OF METALS; MASS-TRANSFER OPERATION [A L K]

Stroboscope

An instrument for observing moving bodies by making them visible intermittently and thereby giving them the optical illusion of being stationary. A stroboscope may operate by illuminating the object with brilliant flashes of light or by imposing an intermittent shutter between the viewer and the object The rate and duration of the visible periods are adjustable

Stroboscopes are used to measure the speed of rotation or frequency of vibration of a mechanical part or system They have the advantage over other instruments of not loading or disturbing the equipment under test Mechanical equipment may be observed under actual operating conditions with the aid of stroboscopes Parasitic oscillations, flaws, and unwanted distortion at high speeds are readily detected It is more economical to obtain this information with a stroboscope than with high-speed motion pictures, and the results are immediately available. Stroboscopes have been employed with balancing machines to locate the lack of balance in lightweight rotating equipment.

By adjusting the rate of viewing to coincide with a multiple of the rate the moving object returns to the same position, parts with a limited range of motion, as in rotatio or vibration, may be

made to appear stationary. The object under view is seen only when it is in one position in space and is freed of the blur normally associated with motion. If the viewing rate is slightly greater than the repetition rate of the motion being observed, the object will appear to move slowly in reverse direction. Similarly, a slightly slower viewing rate makes the object appear to move slowly in the direction of actual motion

The flashing light stroboscopes employ gas discharge tubes to provide a brilliant light source of very short duration A brilliant source is required because of the short period of illumination. Background illumination should be kept as low as possible Tubes may vary from neon glow lamps, when very little light output is required, to special stroboscope tubes capable of producing flashes of several hundred thousand candlepower with a duration of only a few millionths of a second

Tubes may be fired by voltage peaking transformers or by *electronic control circuits.* When *peaking* transformers are employed, the flashes are synchronized with the supply line. Electronic *controls provide adjustable* or synchronous flashing frequencies. Calibrating circuits give direct indication of the frequency of the flashes [A.R.E.]

Stroboscopic photography

Stroboscopic or "strobe" photography is generally understood to refer to pictures of both single and multiple exposure taken by flashes of light from electrical discharges Originally the term referred to multiple-exposed photographs made with a stroboscopic disk as a shutter. One essential feature of modern stroboscopic photography is a short exposure time, usually much shorter than can be obtained by a mechanical shutter.

High-speed photography with stroboscopic light has proved to be one of the most powerful research tools for observing fast motions in engineering and in science Likewise the electrical system of producing flashes of light in xenon-filled flash lamps is of great utility for *studio, candid*, and press photography Spark photography, especially with short air gaps at high voltage, has been used for many years to take short-exposure photographs

Energy storage. Devices for storing energy are required to produce the high peak power needed for producing pictures with a short exposure time. The electrical capacitor is ideal for this service. The energy stored in a capacitor is

$$\frac{CE^2}{2} \quad \text{watt seconds (or joules)}$$

where C is the capacitance in farads and E is the initial voltage to which the capacitor is charged. The electrolytic type of capacitor has the largest ratio of stored watt seconds per pound, but is limited to about 500 volts. Paper capacitors are in widespread use for short-duration flash lights where high voltage is required or where the flash cycle is repeated frequently

Flash lamps. A gaseous discharge lamp or spark gap is required to convert the stored energy of the capacitor into light. The open spark in air is used where a small volume source of very short duration is required. Otherwise the more efficient xenon-filled flash lamp is in almost universal application since its efficiency is about five times greater than that of the open spark. There is an afterglow in xenon gas of about 10 μsec or more which is objectionable when fast subjects such as high-speed bullets are photographed.

Flash duration. Light duration from a xenon lamp or a spark gap is influenced by the electrical characteristics of the *capacitor, the series inductance* of the capacitor and leads, and the *resistance* of the leads. For air gaps, the effective resistance of the gap is negligible compared to these other factors However, a *long xenon-filled flash lamp* of small diameter has a high resistance that limits the current flow. The flash duration of a xenon lamp can be much longer than for an air spark.

Circuit of a flash lamp. Figure 1 shows an electronic flash lamp L such as a xenon-filled glass tube with one external and two internal *electrodes*. The lamp is connected directly across the main flash capacitor C which stores the energy that is to be converted by the lamp into light. Current *from the* transformer, after being rectified, charges the *capacitor* C to a voltage equal to the peak voltage of the secondary of the transformer T.

An electrical discharge is started *in the* flash lamp when switch S is closed An X contact in a synchronized camera shutter can serve as the switch *S. The pulse voltage* in the secondary of the step-up transformer T_2 is connected to the *external* electrode on the lamp and starts a current *in the tube by condenser* action Once the glow on the tube walls is st[illegible]

brilliantly by [illegible]

transient curr[illegible]

to a peak and [illegible]

[illegible] CON-

DUCTION IN GASES.

The complete history of a discharge, showing light as well as current and voltage against time, must be measured experimentally for the exact

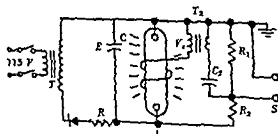


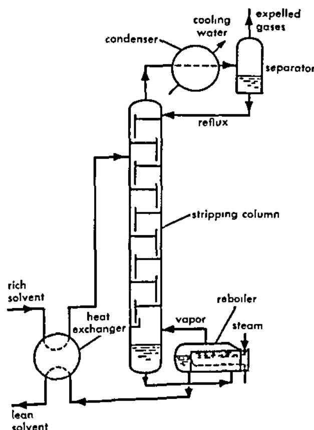
Fig 1. Elementary circuit of an electronic flash lamp L, with main flash capacitor C and trigger circuit with transformer T_2 The lamp L flashes when the synchronizing switch S is closed.

groups being wholly carnivorous or insectivorous. All owls lay the white eggs typical of hole-nesting birds, but a few species nest on the ground or in old hawk or crow nests. Incubation is usually by the female which is the larger member of the pair. *See* AVES. [K.C.P.]

Stripping

The removal of a volatile component from a liquid by vaporization. The stripping operation is an important step in many industrial processes which employ absorption to purify gases and to recover valuable components from the vapor phase (*see* GAS ABSORPTION OPERATIONS). In such processes, the rich solution from the absorption step must be stripped in order to permit recovery of the absorbed solute and recycle of the solvent.

The stripping of a volatile component from a liquid may be accomplished by pressure reduction, the application of heat, or the use of an inert gas (stripping vapor). Many processes employ a combination of all three; that is, after absorption at elevated pressure, the solvent is flashed to atmospheric pressure, heated, and admitted into a stripping column which is provided with a bottom heater (reboiler). Solvent vapor generated in the reboiler or inert gas injected at the bottom of the column serves as stripping vapor which rises countercurrent to the downflowing solvent. When steam is used as stripping vapor for a system which is not miscible with water, the process is called steam stripping *See* DISTILLATION.



Schematic diagram of stripping column and accessory equipment.

Equipment used for stripping operations resembles that employed for absorption and distillation and consists generally of countercurrent columns of the bubble-plate or packed types. The rich solvent is admitted near the top so that the major portion of the column is active for stripping, however in many cases, a short section is provided above the feed point and condensate is refluxed to this section to minimize losses of solvent or to return volatile components which are to be retained in the liquid phase. Stripping columns are also referred to as strippers, desorbers, regenerators, reactivators, and stills (if heat is used).

The operation of stripping is essentially the reverse of absorption in that mass transfer occurs from the liquid to the gas phase. This requires that the partial pressure of solute in the gas be less than the equilibrium vapor pressure of solute over the solution (or, with reference to the liquid, the concentration of solute in the liquid must be greater than the concentration in equilibrium with the gas phase). The stripping operation is also closely related to distillation and in some instances, such as the stripping of hydrocarbons from an absorption oil, the process can be considered the equivalent of a distillation operation in which the bottom and top products have widely different boiling points.

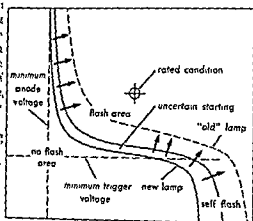
It should be noted that the term stripping has other technical meanings besides that discussed above, including the removal of organic or metal coatings from solid surfaces and the removal of color from dyed fabrics. *See* DYEING; ELECTROPLATING OF METALS; MASS-TRANSFER OPERATION [A.L.K.]

Stroboscope

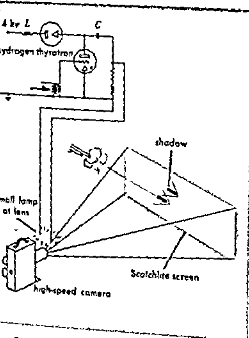
An instrument for observing moving bodies by making them visible intermittently and thereby giving them the optical illusion of being stationary A stroboscope may operate by illuminating the object with brilliant flashes of light or by imposing an intermittent shutter between the viewer and the object The rate and duration of the visible periods are adjustable.

Stroboscopes are used to measure the speed of rotation or frequency of vibration of a mechanical part or system. They have the advantage over other instruments of not loading or disturbing the equipment under test. Mechanical equipment may be observed under actual operating conditions with the aid of stroboscopes. Parasitic oscillations, flaws, and unwanted distortion at high speeds are readily detected. It is more economical to obtain this information with a stroboscope than with high-speed motion pictures, and the results are immediately available. Stroboscopes have been employed with balancing machines to locate the lack of balance in lightweight rotating equipment.

By adjusting the rate of viewing to coincide with a multiple of the rate the moving object returns to the same position, parts with a limited range of motion, as in rotation or vibration, may be



...ig 2 Idealized starting characteristic of an elec...ronic flash lamp showing regions of no flash, flash, ...nd self-start limit as a function of anode voltage and ...igger voltage



...g 3 An effective method of taking a series of pic...ures at high frequency of a bullet in flight against ... Scotchlite screen A small stroboscope lamp is ...ounted as near to the lens as possible The discharge ...apacitor C is charged to 8 kv through the inductance ... from the 4-kv power supply Discharge of the hy...rogen thyratron starts the lamp and a pulse of light ...esults. The hydrogen thyratron has a rapid deioniza...on time and therefore the lamp can be flashed at ...gh frequency such as 10 kc. A Scotchlite screen is ...ne which reflects light from a source back to the ...ource The light received at the film can be several ...undred times more than that received from a flat ...hite screen

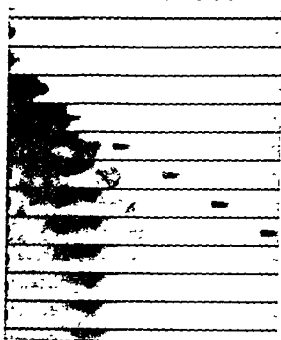


Fig 4. Photographs of a bullet in flight at a velocity of 2000 ft/sec Equipment similar to that described in Fig 3 was used to obtain this series (Harold E. Edgerton)

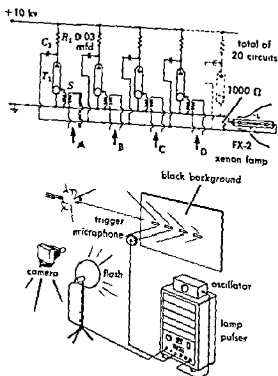


Fig 5 Circuit for flashing a series of capacitors C into a single flash lamp FX-2 at rates up to 100 kc, with 1 5 watt-sec/flash A series of triggering pulses is sent through a gating circuit into terminals A, B, C, etc. The mercury control tube, T_1, deionizes rapidly.

conditions of flash. Data from such tests are available in graphic or tabular form for some of the flash lamps now in use.

Electrical characteristics. As has been described, the current in a long-gap xenon tube rises rapidly from zero after the lamp has been triggered to a peak and then decays in an exponential manner to zero. For many lamps of the xenon-filled types of long arc length, the instantaneous ratio of lamp voltage to current is essentially constant over most of this exponential-like discharge.

Lamp resistance R can be defined, then, as the ratio of the initial anode voltage to the peak current. It follows that the time constant of the current decay is RC seconds and the time constant of the flash duration is $RC/2$ since the light is approximately proportional to the power (current × voltage). These approximate relationships are of great assistance in calculating the expected performance of proposed systems.

As an example, the standard flash lamp type FX-1 has a resistance of $R = 2$ ohms when flashed with 100 μf at 2000 volts. This lamp has an arc length of 6 in. and a 4-mm inside diameter. The resistance of other lamps can be estimated by assuming that a lamp resistance increases directly with length and inversely with the area if the energy density and initial voltage gradient are the same in the two lamps. Thus

$$\text{Flash duration} = \frac{RC}{2} = 2\,\frac{l}{6}\,\frac{4^2}{d^2}\,\frac{C}{2} \quad \text{sec}$$

where l is tube length in inches and d is inside diameter in mm. This approximate equation holds for a flash tube with 20-cm filling of xenon and with an energy density of at least 40 watt-sec/cm^3.

As previously mentioned, spark gaps in air do not follow this rule because their resistance is usually so small that the circuit inductances form the current-limiting impedances. The discharge current from a capacitor into an air spark gap is oscillatory, at a frequency set by the circuit constants. A pulse of light follows in general the major envelope of the current oscillations. Afterglow in the excited gas in the gap produces light between half cycles of current.

Time between flashes. A flash lamp in a circuit such as Fig. 1 cannot produce a second flash of light until the capacitor has been recharged. When a battery or other dc supply is used instead of the rectifier-transformer arrangement of Fig 1, the charging rate is the straightforward RC time constant of the circuit elements In the interval comprising four time constants the voltage will be up to 98% of the final value and the light about 96% For any specific example, the charging time can be made shorter by decreasing the series charging resistance

As this resistance is reduced for faster charging, a new condition called holdover is encountered in which the flash lamp does not extinguish but is continuously excited by the charging current. The minimum value of the holdover current is about 1-2 amp for many practical flash lamps

Holdover occurs when the end of the discharge current transient and the charging current are equal. The capacitor is not charged during a holdover condition, and the lamp voltage remains at a relatively low value, with a large current from the dc supply and a large loss of energy and heating in the charging circuit resistor.

Holdover has been experienced in 200-1000 watt-sec flash equipment in the 450- and 900-volt types when 10-sec charging has been desired Some of the flash units are equipped with a relay which inserts a large charging resistor for a few seconds at the beginning of the charging cycle to reduce the current so that the lamp will deionize

Lamps operated with a few watt seconds per flash have been operated at frequencies up to several hundred flashes per second. With less energy per flash, the lamps appear to be able to operate at a higher frequency without skipping or holdover.

Requirements. For each specific use of stroboscopic photography, the user will want to satisfy three requirements: (1) the flash duration required to "stop" the action; (2) the quantity of light adequate to record the data on the film; and (3) the interval of time between flashes for desired displacement-time information.

The first item, required flash duration, is known once the velocity and the blur definition of the subject are available. For example, suppose one wishes to obtain a clear image of a bullet which has a velocity of 2000 ft/sec. Let the definition of the blur on the bullet be less than 0 01 in Now calculate the minimum flash duration from the equation $d = vt$ where d = distance, v = velocity, and t = time Then,

$$t = \frac{d}{v} = \frac{.01}{2000 \times 12} = 0.4\,\mu\text{sec}$$

The second requirement, quantity of light, can best be estimated by the guide factor method as used by photographers. In its simplest form, the guide factor equation is limited to the single-lamp front-lighted case, where the camera lens is at a distance from the subject.

$$DA = \sqrt{(BCPS)\,\frac{S}{C}}$$

where DA = guide factor
D = distance of the lamp to the subject (must be at least 10 reflector diameters)
A = numerical aperture of the lens
$BCPS$ = the beam candle-power second output of the lamp and reflector
S = the ASA exposure index of the film
C = a constant which is 15–25 if D is in feet

The third item, time between flashes, depends upon the information that is desired. If velocity is desired, only two pictures are needed. The required

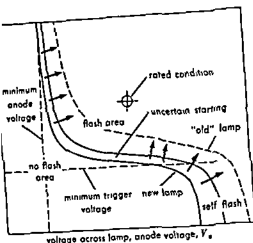


Fig 2 Idealized starting characteristic of an electronic flash lamp showing regions of no flash, flash, and self-start limit as a function of anode voltage and trigger voltage

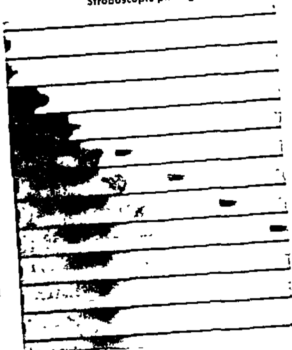


Fig 4. Photographs of a bullet in flight at a velocity of 2000 ft/sec. Equipment similar to that described in Fig 3 was used to obtain this series (*Harold E. Edgerton*)

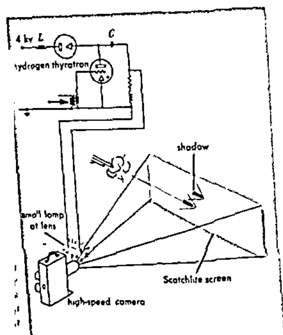
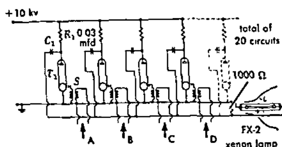


Fig 3 An effective method of taking a series of pictures at high frequency of a bullet in flight against a Scotchlite screen A small stroboscope lamp is mounted as near to the lens as possible The discharge capacitor C is charged to 8 kv through the inductance L from the 4 kv power supply Discharge of the hydrogen thyratron starts the lamp and a pulse of light results The hydrogen thyratron has a rapid deionization time and therefore the lamp can be flashed at high frequency such as 10 kc. A Scotchlite screen is one which reflects light from a source back to the source The light received at the film can be several hundred times more than that received from a flat white screen



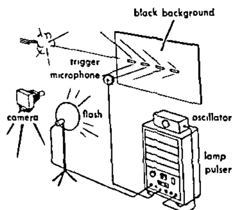


Fig 5 Circuit for flashing a series of capacitors C into a single flash lamp FX-2 at rates up to 100 kc, with 1 5 watt-sec/flash. A series of triggering pulses is sent through a gating circuit into terminals A, B, C, etc The mercury control tube, T_1, deionizes rapidly.

conditions of flash Data from such tests are available in graphic or tabular form for some of the flash lamps now in use.

Electrical characteristics. As has been described, the current in a long-gap xenon tube rises rapidly from zero after the lamp has been triggered to a peak and then decays in an exponential manner to zero. For many lamps of the xenon-filled types of long arc length, the instantaneous ratio of lamp voltage to current is essentially constant over most of this exponential-like discharge.

Lamp resistance R can be defined, then, as the ratio of the initial anode voltage to the peak current. It follows that the time constant of the current decay is RC seconds and the time constant of the flash duration is $RC/2$ since the light is approximately proportional to the power (current × voltage). These approximate relationships are of great assistance in calculating the expected performance of proposed systems.

As an example, the standard flash lamp type FX-1 has a resistance of $R = 2$ ohms when flashed with 100 μf at 2000 volts. This lamp has an arc length of 6 in. and a 4-mm inside diameter The resistance of other lamps can be estimated by assuming that a lamp resistance increases directly with length and inversely with the area if the energy density and initial voltage gradient are the same in the two lamps. Thus

$$\text{Flash duration} = \frac{RC}{2} = 2\,\frac{l}{6}\,\frac{4^2}{d^2}\,\frac{C}{2} \quad \text{sec}$$

where l is tube length in inches and d is inside diameter in mm This approximate equation holds for a flash tube with 20-cm filling of xenon and with an energy density of at least 40 watt-sec/cm^3.

As previously mentioned, spark gaps in air do not follow this rule because their resistance is usually so small that the circuit inductances form the current-limiting impedances The discharge current from a capacitor into an air spark gap is oscillatory, at a frequency set by the circuit constants. A pulse of light follows in general the major envelope of the current oscillations. Afterglow in the excited gas in the gap produces light between half cycles of current

Time between flashes A flash lamp in a circuit such as Fig. 1 cannot produce a second flash of light until the capacitor has been recharged. When a battery or other dc supply is used instead of the rectifier-transformer arrangement of Fig 1, the charging rate is the straightforward RC time constant of the circuit elements. In the interval comprising four time constants the voltage will be up to 98% of the final value and the light about 96%. For any specific example, the charging time can be made shorter by decreasing the series charging resistance.

As this resistance is reduced for faster charging, a new condition called holdover is encountered in which the flash lamp does not extinguish but is continuously excited by the charging current. The minimum value of the holdover current is about 1-2 amp for many practical flash lamps.

Holdover occurs when the end of the discharge current transient and the charging current are equal. The capacitor is not charged during a holdover condition, and the lamp voltage remains at a relatively low value, with a large current from the dc supply and a large loss of energy and heating in the charging circuit resistor.

Holdover has been experienced in 200-1000 watt-sec flash equipment in the 450- and 900-volt types when 10-sec charging has been desired Some of the flash units are equipped with a relay which inserts a large charging resistor for a few seconds at the beginning of the charging cycle to reduce the current so that the lamp will deionize.

Lamps operated with a few watt-seconds per flash have been operated at frequencies up to several hundred flashes per second. With less energy per flash, the lamps appear to be able to operate at a higher frequency without skipping or holdover

Requirements. For each specific use of stroboscopic photography, the user will want to satisfy three requirements: (1) the flash duration required to "stop" the action; (2) the quantity of light adequate to record the data on the film; and (3) the interval of time between flashes for desired displacement-time information

The first item, required flash duration, is known once the velocity and the blur definition of the subject are available. For example, suppose one wishes to obtain a clear image of a bullet which has a velocity of 2000 ft/sec. Let the definition of the blur on the bullet be less than 0.01 in. Now calculate the minimum flash duration from the equation $d = vt$ where d = distance, v = velocity, and t = time. Then,

$$t = \frac{d}{v} = \frac{.01}{2000 \times 12} = 0.4\ \mu\text{sec}$$

The second requirement, quantity of light, can best be estimated by the guide factor method as used by photographers. In its simplest form, the guide factor equation is limited to the single-lamp, front-lighted case, where the camera lens is at a distance from the subject

$$DA = \sqrt{(BCPS)\,\frac{S}{C}}$$

where DA = guide factor
D = distance of the lamp to the subject (must be at least 10 reflector diameters)
A = numerical aperture of the lens
$BCPS$ = the beam-candle-power-second output of the lamp and reflector
S = the ASA exposure index of the film
C = a constant which is 15–25 if D is in fe

The third item, time between flashes, depend upon the information that is desired. If velocity i desired, only two pictures are needed. The require

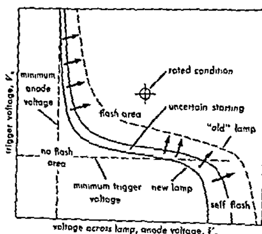


Fig 2 Idealized starting characteristic of an electronic flash lamp showing regions of no flash, flash, and self-start limit as a function of anode voltage and trigger voltage

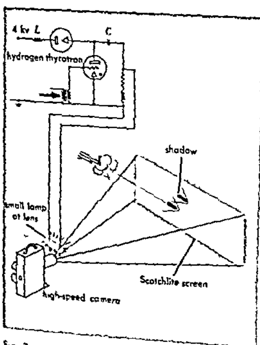


Fig 3 An effective method of taking a series of pictures at high frequency of a bullet in flight against a Scotchlite screen A small stroboscope lamp is mounted as near to the lens as possible The discharge capacitor C is charged to 8 kv through the inductance L from the 4 kv power supply Discharge of the hydrogen thyratron starts the lamp and a pulse of light results The hydrogen thyratron has a rapid deionization time and therefore the lamp can be flashed at high frequency such as 10 kc A Scotchlite screen is one which reflects light from a source back to the source The light received at the film can be several hundred times more than that received from a flat white screen

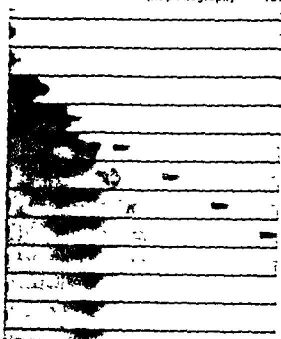


Fig 4 Photographs of a bullet in flight at a velocity of 2000 ft/sec Equipment similar to that described in Fig. 3 was used to obtain this series (Harold E. Edgerton)

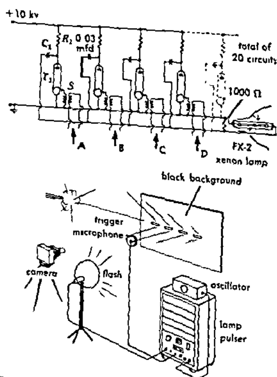


Fig 5 Circuit for flashing a series of capacitors C into a single flash lamp FX-2 at rates up to 100 kc, with 1 5 watt-sec/flash A series of triggering pulses is sent through a gating circuit into terminals A, B, C, etc. The mercury control tube, T_1, deionizes rapidly

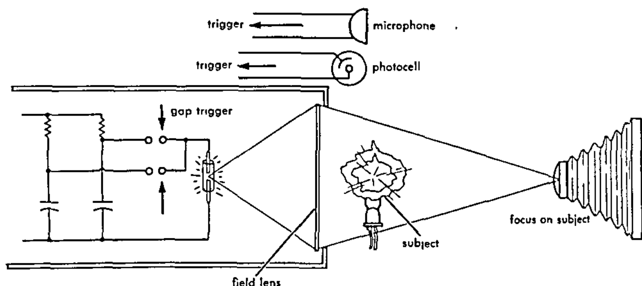


Fig 6 Circuit diagram of a double flash unit where two discharge capacitors are triggered separately by trigger gap circuits A field lens is used to collect the light and direct it into the camera lens. This unit was especially developed to measure the early speed of explosion of a dynamite cap with two superimposed photographs spaced about 5 μsec apart in time

time interval is calculated by the equation that was used for the first requirement. Again using the bullet as an example, the time for the bullet to travel 2 ft is 1 msec. If one wishes to have a sequence of photographs of a bullet striking an object when the bullet motion between exposures is 0 2 ft, then the interval of time between flashes needs to be 100 μsec (at a rate of 10,000 flashes/sec)

The production and control of flashes of light at high frequencies will be discussed later in this article

Conditions for starting. The electrical conditions briefly mentioned before for satisfactory starting of an electronic flash lamp can be presented in chart or curve form such as Fig. 2. Below a minimum trigger voltage and a minimum anode voltage the lamp will not flash, as indicated by the no flash area. With ample starting margin as indicated by the point marked rated condition, the lamp will start reliably every time as desired Now, if the characteristic starting curve engulfs this operating point because of changes during the life of the lamp or circuit, the lamp may begin to misfire at infrequent intervals of time This condition can be corrected by increasing the size of the trigger capacitor C_2 or the initial voltage across C_2, since either of these changes increases the starting voltage V_s.

All of the methods of overcoming missing mentioned at the end of the previous paragraph require an increase of peak trigger current in the switch *S* Should this switch be a delicate X sync contactor in a camera shutter, damage may result to the contacts and the contacts may then fail to close in a positive manner for reliable starting. Furthermore, the contacts usually bounce when they hit, causing sparking at the contacts which reduces the triggering voltage.

A thyratron or a strobotron can be used as a trigger tube in place of the switch *S* to overcome these difficulties. Then the lamp can be triggered by a microphone, photocell, or other electrical pulse signal.

High-frequency flashes. The frequency limitations discussed previously can be avoided by several electrical circuits which do not depend upon the holdover characteristics of the flash lamp

1 A control tube, such as a hydrogen thyratron, a hydrogen gap, or a mercury pool tube can be used to switch the current into the flash lamp These tubes have a rapid deionization of the gas that is used to conduct the pulses of current Equipment has been developed which operates in short bursts up to many thousands of flashes per second (Figs 3, 4)

2 A series of separate capacitors which are switched into a flash lamp by means of switching tubes is capable of operating at 10-μsec intervals with 1 5 watt-sec/pulse (Fig. 5)

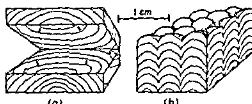
3 Separate flash lamps, each with a complete flashing circuit, can be used Some arrangement is required to trigger the lamps at the desired intervals of time (Fig 6).

See PHOTOGRAPHY, SPARK, ELECTRIC; STROBOSCOPE [H E.E.]

Bibliography Bibliography on high-speed photography *Soc Motion Picture Television Engrs.* 56:93–111, 1951, W D Chesterman, *The Photographic Study of Rapid Events*, 1951; R. B Collins (ed) *High-Speed Photography*, 1956–1957; H E. Edgerton and J. R Killian, Jr., *FLASH! Seeing the Unseen by Ultrahigh-Speed Photography*, 2d ed., 1954; G. A Jones, *High-Speed Photography*, 1953

Stromatolite

A structure in calcareous rocks consisting of concentrically laminated masses of calcium carbonate and calcium-magnesium carbonate which are believed to be of calcareous algal origin. These structures are irregular to columnar and hemispheroidal



Sectioned stromatolites. (a) Diagram of part of a large, strongly laminated stromatolite, showing how the laminae dip away from the center of the hemispheroidal mass if viewed from above, and toward the center if viewed from beneath (b) A small stromatolite consisting of columns with highly arched laminae Lower Ordovician of western Wisconsin (From R R Shrock, Sequence in Layered Rocks, McGraw-Hill, 1948)

in shape and range from 1 millimeter to many meters in thickness They may be small buttons or biscuits or may have areal extents as great as 1 or more square kilometers Stromatolites are found in rocks ranging from Precambrian to Recent age. See ALGAE FOSSILS.

Stromatolite structures, developed in growth position, have convex surfaces upward These struc-tures are of . . .

. . . component of, the marine ecologic community of the shallow seas on sedimentary shelves and platforms See BIOHERM, BIOSTROME; ORGANIC REEF.

[S A.W.E.]

Bibliography P E Cloud, Jr., Notes on stromatolites, *Am J Sci.*, 240.363–379, 1942, F. J Pettijohn, *Sedimentary Rocks*, 2d ed., 1957; R. R Shrock, *Sequence in Layered Rocks*, 1948.

Stromatoporoidea

An extinct order of conspicuously layered, calcareous, animal skeletons belonging either to the class Hydrozoa or Scyphozoa The fossils occur as incrusting or massive colonies, usually attached to corals or other animals, in marine rocks of Ordovician, Silurian, and Devonian age During this time they made beds of limestone and built hundreds of reefs which were buried by sediments Petroleum occurs in some vertical structures formed by stromatoporoids Stromatoporoids have been classified in various larger groups by different authors In features of organization they are higher than sponges and lower than corals, medusae are not known The consensus by specialists who study them is that they should be ranged among the hydrozoans See H...

Most sk-
and range
right cyli... s or bushes, ranging up to 10 cm in diameter and up to 3 m in length A characteristic feature is the development of annual layers, or latilaminae These are 2–10 cm thick and fairly regularly arranged (Fig 1) The surface is either smooth or has round or elongate elevations. Many species have radiating and branching grooves on the surface of the laminae. These furrows, called astrorhizae because they generally are arranged in a starlike pattern, may have housed reproductive polyps (Fig. 2).

The identification of stromatoporoids requires thin sections cut vertical and tangential to the laminae. Internally the skeleton consists of curved plates, with or without vertical pillars, in the oldest, most primitive forms (Fig. 3). Most of the skeleton consists of thin plates (laminae) 0.01–0.1 mm thick, with short or long vertical pillars, and columns of upturned laminae and thicker pillars (Fig. 4). The finer tissue, which is important in



Fig. 1 Annual layers of stromatoporoid *Clathrodictyon* Silurian

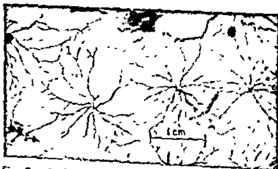


Fig 2. Surface of stromatoporoid *Ferestromatopora* showing astrorhizae. Silurian.



Fig 3 Vertical section of primitive stromatoporoid *Labechia* showing curved plates and long pillars. Ordovician

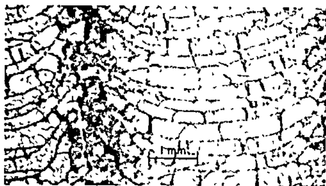


Fig 4. Vertical section of Devonian stromatoporoid *Anostylostroma* showing short pillars, columns of upturned laminae, and thicker pillars.

generic and specific identification, is either homogeneous, flocculent, transversely fibrous and porous, or is full of light and dark dots called maculae.

The Stromatoporoidea, consisting of 5 families and 35 genera, flourished in Ordovician and Silurian time, reached their acme of complexity in structure and numbers of individuals in the Devonian, and became extinct at the end of the Devonian. They evolved into the order Sphaeractinoidea of the Upper Paleozoic and Mesozoic, which order in turn gave rise to the modern orders Hydroida and Hydrocorallina. *See* HYDROIDA; SPHAERACTINOIDEA. [J.J.G.]

Bibliography: J. J. Galloway, Structure and classification of the Stromatoporoidea, *Bull Am. Paleontol.*, 37(164):345–470, 1957; H. A. Nicholson, *A Monograph of the British Stromatoporoids*, Palaeontol. Soc., 1886–1892.

Strongyloidea

Roundworms which as adults usually inhabit the gastrointestinal tract, kidney, or respiratory tract of amphibians, reptiles, birds, and a wide variety of mammals including man. Taxonomically, they are considered by some authorities to constitute an order, by others, a superfamily of the class Nematoda

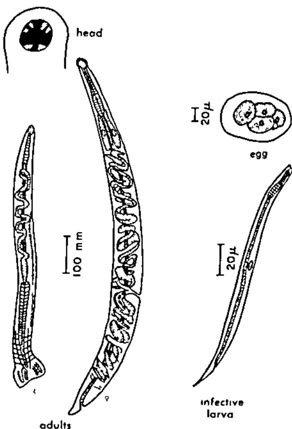
Morphology and life history. The male worms are characterized by a caudal cuticular bursa copulatrix supported by rays The esophagus is usually more or less club-shaped posteriorly, but without a definite spherical bulb and without a valvular apparatus. They are usually oviparous, occasionally viviparous Each female hookworm of man passes from 5000–20,000 eggs per day. Larval stages of Strongyloidea are usually passed in the soil but they are found rarely as parasites of earthworms and snails.

Infection occurs through the penetration of the host's skin by the larvae, through the ingestion of larvae in water or on grass, or by the ingestion of larvae in an intermediate host harboring the larvae. The larvae migrate to their adult habitats directly through the tissues or by way of the bloodstream. The adults live varying lengths of time, up to 14 years

Economic importance. Some of the most important parasites of man and his domestic animals belong to this group and the morbidity, mortality and economic loss due to them are extensive The hookworms of man, *Ancylostoma* and *Necator* (see illustration), are estimated to infect more than 456,800,000 persons in the tropical and subtropical areas of the world. They live attached to man's small intestine and suck blood, producing varying degrees of anemia. Horses, cattle, sheep, swine, and other domestic and wild animals are hosts to a large number of Strongyloidea which cause damage by sucking blood and through tissue inflammation and destruction. The migration of larval and maturing worms through the various organs may cause widespread damage and hemorrhage. In contrast to specific and fatal diseases caused by bacteria, viruses, and protozoa, the damage caused by these worms is usually not spectacular but is insidious, resulting in anemia, loss of weight, failure to gain weight, greatly reduced efficiency, and emaciation.

The damage to the host depends upon the species and numbers of worms and the age and resistance of the host Thus, 500–1000 hookworms can cause a severe anemia in man, yet horses may harbor from 30,000 to 50,000 strongyles without recognizable symptoms Young animals are not as resistant to worm infections as are mature animals. Worm parasites of other animals occasionally infect man

The larval forms of many of the domestic animal Strongyloidea are quite resistant to heat, cold, and



A hookworm, *Ancylostoma duodenale*

mon Strongyloidea

Family and genus	Host	Length of adult female worm, mm	Disease
Trichostrongylidae			
Cooperia	Cattle, sheep	6	Failure to gain weight
Haemonchus	Cattle, sheep, goat	30	Wireworm anemia
Mecistocirrus	Cattle, sheep, swine	29	Anemia
Nematodirus	Sheep, cattle, goat	20	Irritation of intestine
Ollulanus	Cat	1	Irritation of stomach
Ostertagia	Cattle, sheep	9	Stomach inflammation, diarrhea
Trichostrongylus	Sheep, cattle	9	Diarrhea, anemia
Ancylostomidae			
Ancylostoma	Man, dog, cat	11	Hookworm anemia, creeping eruption
Bunostomum	Cattle, sheep	26	Hookworm anemia
Necator	Man, swine	11	Hookworm anemia
Uncinaria	Dog, cat, fox, swine	8	Hookworm anemia
Strongylidae			
Oesophagostomum	Cattle, sheep, swine	15	Nodules in intestine, ulceration
Stephanurus	Swine	40	Liver, kidney, lung damage
Strongylus	Horse	55	Anemia, organ damage, intestinal irritation
Syngamidae			
Syngamus	Birds, chicken, turkey, mammals	40	"Gapes," blockage of trachea, anemia
Metastrongylidae			
Crenosoma	Fox, cat, dog	13	Inflammation of blood vessels
Dictyocaulus	Sheep horse, cattle	100	Pulmonary irritation
Metastrongylus	Swine	12	Pulmonary irritation
Diaphanocephalidae			
Kalicephalus	Snakes	30	
Pseudaliidae			
Torynurus	Whales, porpoises	45	Inflammation of bronchi and blood vessels
Heligmosomidae			
Nippostrongylus	Rodents	6	Emaciation

ryness and therefore their geographical distribution is extensive and not limited to areas with warm moist climate as is hookworm

Diagnosis and treatment. The diagnosis of parasitic worms cannot be made clinically but depends upon the detection of eggs or larvae in the stool, urine, or sputum Unfortunately since there are many varieties of roundworms parasitic to the gastrointestinal tract in domestic animals and since many of their eggs are similar, specific diagnosis by this means is unsatisfactory Positive diagnosis is made at autopsy Fortunately, man harbors relatively few species of roundworms, and the egg of hookworm his only *strongyloid parasite* can easily be identified A number of chemotherapeutic agents notably carbon disulfide, copper sulfate, phenothiazine, and tetrachlorethylene, are effective against several of the Strongyloidea but for some species there is no effective therapy The presence of worms produces varying degrees of immunity in the host but no substance immunizing against infection is available.

Prevention. Prevention of infection in man is easy in theory and consists of sanitary disposal of human excrement in a simple pit privy In practice, the use of night soil for crop fertilizer, religious customs, and ignorance continue the 2,000,000 000 helminth infections in man throughout the world The control of infection in domestic animals is dependent upon proper handling of their manure and rotation of pastures. Chemotherapy is helpful to the individual but is valuable in control only when used in conjunction with sanitary measures and, in the case of man, education *See* HOOKWORM; HOOKWORM DISEASE; NEMATODA.

[H.W BR]

Strongyloidiasis

An infestation of man with one of the roundworms of the genus *Strongyloides* The presence of *Strongyloides stercoralis*, a minute nematode, in the small *intestine or lungs, interferes* with the mucosae and is capable of producing a catarrh. While only the female occurs in the intestine, both

sexes may exist in the lungs. The worm lays eggs that hatch in situ; larvae pass out in feces and sometimes sputum They may reinvade the tissues to complete their life cycle; this is known as autoinfection or hyperinfection. Those larvae in soil transform into free-living adults, from which other free generations arise, as well as infective larvae which may enter through the skin Their biology and epidemiology parallel hookworm disease. Gentian violet is the only, partially successful, therapeutic agent. *See* PARASITOLOGY, MEDICAL; STRONGYLOIDEA. [J.F.MA.]

Bibliography: C. F. Craig and E. C. Faust, *Clinical Parasitology*, 5th ed., 1951.

Strontianite

The mineral form of strontium carbonate, usually with some calcium replacing strontium. It characteristically occurs in veins with barite or celestite, or as masses in certain sedimentary rocks Strontianite has orthorhombic symmetry and the same structure as aragonite It is normally prismatic with the development of pseudohexagonal form, but it may also be massive. It may be colorless or gray with yellow, green, or brownish tints The hardness is 3½, and the specific gravity 3.76

It occurs at Strontian, Scotland. and in Germany, Austria, Mexico, and India and, in the United States, in the Strontium Hills of California. *See* CARBONATE MINERALS; STRONTIUM [R.I.H.]

Strontium

Chemical element number 38, strontium, Sr, is the least abundant of the alkaline-earth metals. It occurs to the extent of 0.042% in the earth's crust, making it approximately as abundant as the elements sulfur and chlorine. The main ores include celestite, $SrSO_4$, and strontianite, $SrCO_3$, which



are found chiefly in Scotland, Arkansas, and Arizona. Strontianite is colorless to light greenish or reddish, depending upon the impurities The major contaminants of these ores are iron, aluminum, and calcium. The fact that calcium and strontium ions are close enough in size to replace one another in their compounds explains both the large calcium impurity in strontium ores and the uptake of the radioactive fission product Sr^{90} in the calcareous parts of plants and animals

It was Adair Crawford who, in 1790, first distinguished between naturally occurring strontium and barium carbonates, using a sample which had been unearthed in a lead mine in Strontian, Scotland. This was confirmed in 1792 by Thomas Hope. The name of the carbonate ore and the element itself stem from the location of the original discovery

Uses. Strontium nitrate is used in pyrotechnics, railroad flares, and tracer bullet formulations. Strontium hydroxide forms soaps and greases with a number of organic acids which are structurally stable, resistant to oxidation and breakdown over a wide temperature range, and resistant to disintegration by water and the leaching action of hydrocarbons. Other strontium compounds are used as paint driers and have minor medical uses

Extraction of the metal. The pure metal was first isolated by Sir Humphry Davy, who electrolyzed a mixture of strontium and mercuric oxides with a mercury pool cathode; the mercury was then distilled away from the amalgam that had formed, leaving silvery globules of the pure metal The element is produced commercially in the United States on a small scale either by electrolyzing a mixture of potassium and strontium chlorides or by reducing the oxide with aluminum in a vacuum at such a temperature that the strontium distills out

The metal is silvery-white and lustrous, and it quickly forms a protective oxide coating in the air It is softer than calcium, and vigorously dissolves in water and acids to yield hydrogen gas. The element combines readily with oxygen to form the oxide when heated, but will not form the nitride unless heated above 380°C The physical properties of the element are given in the table.

Properties of strontium

Atomic number	38
Atomic weight	87.63
Isotopes (stable)	84, 86, 87, 88
Electron configuration	2 8 18 8 2
Ionic radius, A	1.13
Boiling point, °C	1638 (?)
Melting point, °C	704 (?)
Atomic volume, cm³/g-atom	34.5
Density, g/cm³	2.6
Latent heat of vaporization at boiling point, kilojoules/g-atom	383

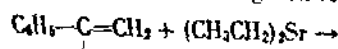
Principal compounds. Celestite is used in the modern process for the production of strontium compounds in the United States, and is obtained in two grades, 92% and 84–92% strontium sulfate The ore is first treated with 10% hydrochloric acid and then with water, which leaches out the calcium sulfate and carbonate that are the chief impurities A small excess of sodium carbonate is then added and the solution is shaken at 150–160°F for 6 hours, after which approximately 85% of the strontium is precipitated as the carbonate Part of the supernatant liquid is then removed, more sodium carbonate is added, and the liquid is agitated for 10 more hours The precipi n allowed

settle for 4 hours, after which the clear supernatant liquid is decanted and the precipitate is washed three times with hot water. The precipitate is taken up again in hydrochloric acid and reprecipitated by the use of sodium carbonate. This precipitate is washed several times with hot water until the supernatant liquid is largely chloride-free, and then is filtered, pressed dry, and ground to a powder. The carbonate is sold as a fine white powder of various grades for the preparation of other salts

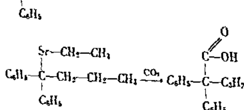
Strontium is divalent in all its compounds which are, aside from the hydroxide, fluoride, and sulfate, quite soluble Strontium perchlorate has an appreciable solubility in organic solvents. The nitride is ionic and reacts with water to give ammonia, whereas strontium carbide, SrC_2, releases acetylene upon hydrolysis The sulfide may be formed by reduction of the sulfate with charcoal, and it undergoes hydrolysis in water to form $Sr(SH)_2$ Strontium is a weaker complex former than calcium, giving a few weak oxy complexes with tartrates, citrates, and so on. The large electrical conductivity of strontium boride, SrB_6, approaches values which are characteristic of metals

The reaction of strontium with hydrogen is vigorous and complete at temperatures of a few hundred degrees The hydride, SrH_2, reacts with alcohols to form the alcoholates and may be used in place of calcium hydride for organic condensation and reduction reactions The complex alkyl, S-Z-(CH ...

lution, behaves as a Grignard reagent



$$C_6H_5{-}C(C_6H_5){=}CH_2 + (CH_3CH_2)_2Sr \rightarrow$$



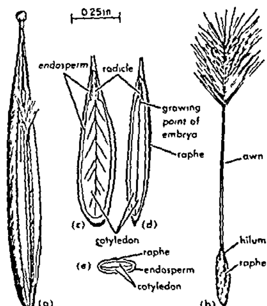
$$C_6H_5{-}C(Sr{-}CH_2{-}CH_3)(C_6H_5){-}CH_2{-}CH_2{-}CH_3 \xrightarrow{CO_2} C_6H_5{-}C(COOH)(C_6H_5){-}C_3H_7$$

Analytical methods. The determination of strontium involves the prior separation of any barium by precipitation as the insoluble chromate and the subsequent precipitation and weighing of strontium sulfate Strontium salts can be recognized qualitatively by their vivid crimson flame coloration See ALKALINE-EARTH METALS [R F R]

Strophanthus

A genus of woody climbers of the dogbane family (Apocynaceae) natives of tropical Asia and Africa They are the source of arrow poisons The dried greenish, ripe seeds of *Strophantus hispidus* and *S kombe* contain the glucoside, strophanthin, which is much used in treating heart ailments. Strophanthin acts directly on heart muscle, increasing muscular force It causes the heart to beat more regularly and decreases the pulse rate Strophanthin is also a precursor of cortisone used in the treatment of arthritis. See GENTIANALES.

[P.D S.]



Seed of *Strophanthus kombe*. (a) Follicle. (b) Seed, natural size. (c, d, e) Sections of the seed. (From *T. E. Wallis, Textbook of Pharmacognosy, 3d ed., Little, Brown, 1955*)

Structural analysis

All structures must be designed to carry loads without danger of over-all collapse or failure of their components. One way to assure structural safety is to ascertain that stresses and strains produced by loads are less than those allowed by established design codes. The determination of stresses and strains in proposed new structures is a primary objective of structural analysis See STRESS AND STRAIN.

Generally, analysis begins with a check of the over-all stability of a structure. This involves first the determination of the forces exerted by the structure against its supports. If the supports are adequate to withstand these forces, they, in turn, react with equal but opposite forces against the structure

If computation shows that the reactions balance the loads (weight of structure, occupants, stored materials, vehicles, wind, earthquake forces), the structure is in static equilibrium See STATICS

The next step is determination of internal forces and unit stresses in the components of the structure. Finally, if necessary, the deformation of the structure as a whole and of its components may be calculated (see STRUCTURAL DEFLECTIONS). These steps are facilitated by use of principles and equations, such as the law of equilibrium, method of least work, moment distribution and dummy unit-load method Many of these tools are based on the assumption that the structure is elastic under loads, that is, stress is proportional to strain.

Basic principles. The law of equilibrium is basic in structural analysis. It is useful in computing reactions of beams, trusses, frames, arches and other structures, as well as stresses. For example, the structures in Fig. 1 are acted upon by coplanar, nonconcurrent force systems, or loads and reactions. These must balance if the structures are stable.

When in equilibrium, the structures must satisfy three conditions, or equations: (1) the sum of the horizontal components of all the forces must equal zero; (2) the sum of the vertical components must equal zero; (3) the sum of the moments about any axis normal to the plane must equal zero.

The three independent equations permit three unknowns to be found. Thus, the equilibrium equations can be used directly to compute the reactions, bending moments, shears, and therefore the stresses When these equations are satisfied, a structure is said to be statically determinate. Figure 1*a* is an example of such a structure; there are three unknowns—vertical and horizontal components of the reaction at the left end and a vertical reaction at the right end.

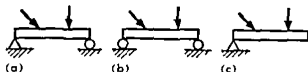


Fig. 1 Stable and unstable structures. The structure in (a) is stable, those in (b) and (c) are not The triangle indicates that the structure is attached, the circle that the structure, while supported, can move

The structures in Fig. 1b, with no horizontal reaction to balance the horizontal load, and Fig 1*c*, with no balancing support moment, do not satisfy the law of equilibrium; they are unstable

Another basic tool of structural analysis is the principle of superposition. It states that the total moments, shears, stresses, and deflections caused by a group of loads are equal to the sum of the effects of the separate loads, if the combined effects do not stress the material beyond the elastic range

Statically indeterminate structures. When the number of reaction components exceeds the number of independent equations that must be satisfied by the loads and reactions when equilibrium exists, the structure is statically indeterminate For example, if the columns of the rigid frame in Fig. 2*a* are fixed at their bases d and e, there will be six reaction components, as indicated in Fig 2*b*. This structure will be indeterminate to the third degree, there are six unknowns, whereas the law of equilibrium yields only three equations

Approximate methods are available for the analysis of statically indeterminate structures These methods are based on the results of exact analysis or examination of a model constructed of flexible splines. Such studies show that in a rigid frame restrained against rotation at the base, the members change curvature somewhere near their mid points, as indicated in Fig. 2*c*. Bending moment is zero at such points. Thus, these points of contraflexure are equivalent to pin joints.

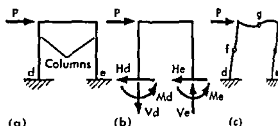


Fig. 2. (a) A statically indeterminate structure. (b) H_d and H_e are the horizontal reactions to P at points d and e, V_d and V_e the vertical reactions, and M_d and M_e the moments. (c) The points of contraflexure or changes of curvature are f, g, and h.

If their locations are assumed in the columns and the girder of the frame in Fig. 2, and if it is also assumed that column shears are each equal to one half the lateral load, the reactions may be found from the equations of equilibrium Similar approximate methods (described in the following subsection) have been used to estimate wind stresses in *building frames* (Fig. 3).

Approximate methods. The portal method of computing wind stresses in multistory frames assumes that there are points of contraflexure at the midpoints of columns and girders and each interior column will receive twice as much shear as an exterior column. The resulting structure is statically determinate.

The cantilever method also assumes that *there are points of contraflexure at the midpoints* of columns and girders In addition, it assumes that the unit direct stresses in the columns vary as the distances of the columns from the center of gravity of the frame. Again a statically determinate structure results

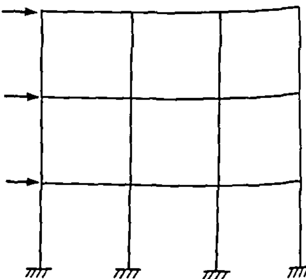


Fig. 3 Diagrammatic representation of the frame of a building showing approximate method of estimating wind stresses (as indicated by ar ws).

(a) (b)

Fig. 4. (a, b) Moments at ends of members in a structure.

Some approximate methods for finding the effects of horizontal loads on multistory frames have been proposed that take into account the stiffness of the members. These include the Witmer K-percentage method, the factor method, and the C method.

Many high building frames were designed with the aid of the portal and cantilever methods at a time when the only exact methods available were the laborious, time-consuming use of Castigliano's theorem and the method of least work. Now, with additional exact methods (slope deflection and moment distribution) and computing machines available, it may be that the approximate methods are obsolete.

Castigliano's theorem may be applied to continuous beams and to frames of the type shown in Fig. 3. It states that the first partial derivative of the strain energy with respect to any particular force (or moment) is equal to the displacement (or rotation) of the point of application of that force (or moment) in the line of its application.

This theorem is closely related to the least work principle, which says that in a loaded statically indeterminate structure the total internal work is the minimum consistent with equilibrium. For example, consider Fig. 2 and assume that the structure is fixed at the base of the right column. The unknown reactions are H_a, V_a, and M_a. The total work in the structure is expressed in terms of these and the applied load P. Partial derivatives with respect to the three redundant reactions (that is, unknowns) are set equal to zero. These three equations yield the values of the unknowns.

An unsymmetrical frame of the type of Fig. 3 has nine redundants per story, that is, it is indeterminate to the twenty-seventh degree. Its solution by the method of least work is cumbersome.

In setting up the work equation for direct stress and moment, respectively, use is made of the following terms

$$W = \frac{S^2L}{2AE} \quad \text{and} \quad W = \frac{M^2\,dx}{2EI}$$

where W is work, S the direct stress in a member, L the length of the member, A its cross section, E the modulus of el...

...ect stress is small and may be neglected.

Slope-deflection method. This method is based on the equation

$$M_{ab} = \frac{2EI}{L}(2\theta_a + \theta_b - 3R) \pm M_{F_{ab}}$$

where M_{ab} is the moment at the a end of member ab, E is the modulus of elasticity of the material; I is the moment of inertia of a cross section of ab; L is the length of ab, θ_a is the rotation in radians of the a end of ab; θ_b is the rotation in radians of the b end of ab; R is an angle in radians equal to the movement of b relative to a divided by L; and $M_{F_{ab}}$ is the fixed-end moment that would occur at a if ab were a fixed beam carrying the actual transverse loads on that member.

In Fig. 4 the moments could be written for the ends of the members if θ_d, θ_e, and R (the horizontal movement of de relative to points c and f, divided by the length of ef) were known. Three slope-deflection equations for Fig. 4b may be written to determine these values:

$$M_{dc} - M_{de} = 0 \qquad M_{ed} - M_{ef} = 0$$
$$M_{cd} + M_{dc} + M_{ef} + M_{fe} + P_1L_{cd} = 0$$

These, when solved, yield the unknowns.

Th...

suc...

15...

rot... and an R term obtained from the shear in the columns of each story.

Three-moment equation. The three-moment equation permits analysis of beams that are continuous over more than one span. Such beams are statically indeterminate because there are insufficient equilibrium equations for determination of reactions. They differ from simply supported beams also because they are subjected to bending moments at supports.

The three-moment equation expresses a relation between the moments over three adjacent supports (one or two may be fixed ends). It is derived by equating the slopes of t...

s...

e...

t...

s... three-moment equation is used as many times as there are unknowns.

Moment-distribution method. The moment-distribution method for continuous beams and frames is based on the fact that the moments in the structure in Fig. 5a may be determined by adding the moments from Figs. 5b and c. In Fig. 5b, de is regarded as a fixed beam with a fixed-end moment $M_{F_{de}}$. In Fig. 5c,

(a) (b) (c)

Fig. 5. (a, b, c) Illustrating the moment-distribution method for continuous beams and frames.



Fig. 6. Moment distribution in a continuous beam.

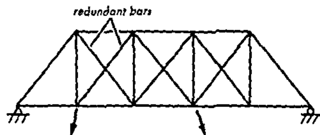


Fig 7. Redundant bars in an indeterminate truss.

the moments at joint *d* will be resisted at *d* by the four members in the ratio of their stiffness (that is, the I/L of each member). Also there will be induced at the far end of each member a moment equal to one-half of the moment at the *d* end, if each member is prismatic.

In the analysis of a continuous beam or frame by moment distribution, the members framing into each joint are considered a fixed-end frame of the type shown in Fig. 5*a* and the moments are distributed as just described. Each such distribution yields an approximate solution. The analyst proceeds from joint to joint, a number of times. For example, for the continuous beam in Fig. 6, the distribution might be started at the center support, then proceed to the supports on both sides, after which the procedure would be repeated. The corrections become smaller and smaller in each cycle

When, as in Figs. 2 or 3, one end of a member moves relative to the other, sidesway occurs. This complicates the solution, but it can be dealt with. Equations can be written in terms of horizontal shears, or converging approximations can be applied, to determine the moments resulting from sidesway.

Indeterminate trusses Trusses may also be indeterminate. They may have redundant bars, that is, bars in excess of the number needed in a statically determinate structure; for example, the double diagonals in the panels of the truss in Fig. 7, if these are made capable of carrying both tension and compression. Like continuous beams, trusses may have redundant reactions; for example, any reaction in Fig. 8*a* (or bar *f*, since its removal will result in a determinate structure).

Types of indeterminate trusses may be solved by the method of least work Also, the truss in Fig. 8 may be analyzed by any method of computing deflections. The deflection at *e* may be calculated with the center support removed as in Fig 8*b*, and the load may be determined that will cause an equal but opposite deflection at *e* as indicated in Fig 8c. The addition of the resulting stresses will be equivalent to those in the truss of Fig. 8*a* in which the deflection at *e* equals zero.

Influence lines. For structures subject to changing or moving load systems, influence lines facilitate analysis. These curves may be plotted for such needed data as reactions, bending moments, shears, deflections, and stresses.

An influence line for bending moment at a point in a beam is a curve that shows the variation of moment at the point as a unit load passes over the structure. For example, M_c in Fig. 9 is the influence line for moment at *c*.

A unit load in the position x as shown in Fig 9 will cause at c a moment of $5x/20$. An ordinate of this amount is plotted under the load As x varies from zero to 15, the moment at c varies from zero to $15/4$. Also, as the load passes from c to a the moment reduces to zero. This influence curve shows that maximum moment at c due to a concentrated load P will occur when the load is at c and will equal $15P/4$. The midpoint of the beam has the highest influence-line ordinate for moment $L/4$ (where L is the span).

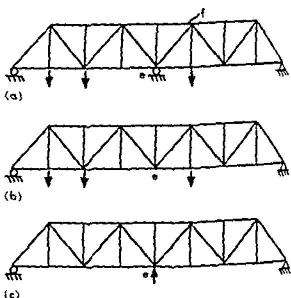


Fig. 8. (a, b, c). Computing deflections in an indeterminate truss.

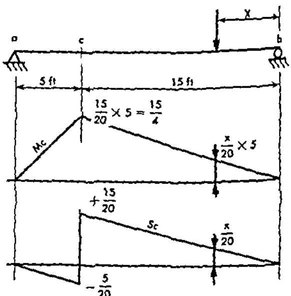


Fig. 9. Influence lines for bending moment M_c at point *c*, and for shear S_c at poin

The influence line for shear at c (S_c in Fig. 9) ows that for maximum shear a single load must placed adjacent to c and on the longer of the gments into which c divides the beam A uniform ad must be placed only on the longer segment for aximum shear. The highest shear influence-line dinate occurs for the point adjacent to the supsrt This ordinate equals one, that is, the shear quals the load

If an indeterminate structure such as that in ig. 8 must be analyzed for moving loads, an influnce line for the reaction at e may be constructed 'lotting this curve can be simplified by applying he principle that a deflection diagram to an approriate scale is an influence line. For the truss in 'ig 8 with the support at e removed, the deflection urve due to a load at e is, to a different scale, the nfluence line for reaction at e.

Model analysis. The relation between deflecion curves and influence lines has led to stress analysis through the use of models Reactions in a continuous beam, such as cde in Fig. 10a may be determined by constructing influence lines with the aid of a spline For example, if deflection is prevented at c and d when the spline is forced out of position a unit distance at e, as indicated in Fig 10b, the curve formed is the influence line for reaction at e Similarly, if the frame in Fig. 2 is restrained against any motion at e, and point d is given a unit horizontal motion (with vertical translation and rotation prevented at d), the distorted frame will be the influence line for H_d, shear at d, when the ordinates are measured normal to the original axis of any member. Similarly, influence lines for axial stress F_d and moment M_d may be constructed

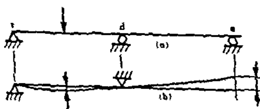


Fig 10 (a) Continuous beam (b) Influence line.

The Beggs method of model analysis utilizes gages that force a model (usually of celluloid) to make small translations or rotation at a cut section At target points that have been set on the model movement is measured with a microscope. The ratio of target movement to movement at the cut section is the same as the ratio of reaction at the cut section to load at the target point. The model need not be composed of prismatic members, they may be tapered, haunched, or curved

Similitude deals with the effects that are introduced because lengths, transverse dimensions, and moduli of elasticity are different in the model and in the actual structure These differences will not alter the influence lines for shear and direct stress. However, influence lines for moment must be multiplied by the ratio of lengths in the actual structure to corresponding lengths in the model. [H L.B.]

Bibliography: M. I. Hetényi, *Handbook of Experimental Stress Analysis*, 1950; J. P. Michalos, *Theory of Structural Analysis and Design*, 1958; H. Sutherland and H. L. Bowman, *Structural Theory*, 4th ed, 1950; S. Timoshenko and D. H. Young, *Theory of Structures*, 1945; *Trans. ASCE*, 107: 925–954, 1942; J. B Wilbur and C. H. Norris, *Elementary Structural Analysis*, 1948.

Structural connections

Means of joining the individual members of a structure to form a complete assembly The connections furnish supporting reactions and transfer loads from one member to another Loads are transferred by rivets, bolts, or welding supplemented by suitable arrangements of plates, angles, or other structural shapes. When the end of a member must be free to rotate, a pinned connection is used.

The suitability of a connection depends on its deformational characteristics as well as its strength Rotational flexibility or complete rigidity must be provided according to the type of structure and degree of end restraint assumed in the design. A rigid connection maintains the original angles between connected members virtually unchanged after loading Flexible or nonrestraining connections permit rotation approximately equal to that at the ends of a simply supported beam. Intermediate degrees of restraint are called semirigid

Framed web connections. A commonly used form of connection for I-beam sections, called a web connection, consists of two angles attached to opposite sides of a member and which are in turn connected to the web of a supporting beam, girder, column, or framing at right angles. A shelf angle may be added to facilitate erection

Riveted or bolted web connections The angles and the rivets that fasten them are designed to transmit shear force only as a simple beam connection. The rotation results from the flexibility of the outstanding angle legs (Fig. 1). For rotation

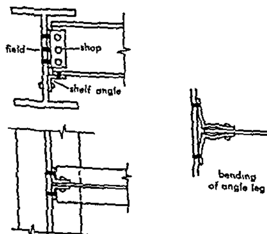


Fig 1. Riveted or bolted web connections

approximating that of a simple beam, the angle legs may become permanently bent. A small end moment is developed by the forces necessary to bend the angles. For building construction, web connections have been standardized for varying beam sizes and are listed in handbooks.

Welded web connections. Flexible welded connections are made with web framing angles attached to the supported beam by fillet welds along the length and ends of the angle legs. These legs are narrow for economy, and the size of weld is such as to resist combined bending and shear without exceeding limiting shear stresses. The legs connected to the supporting member are attached by fillet welds along their outer edges (Fig. 2)

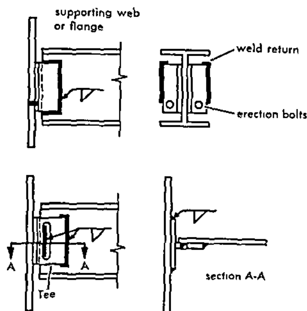


Fig. 2. Welded web connections

To prevent tearing, the edge fillet welds are returned a short distance around the top corners where the stress is greatest. The end rotation of the connected beam results from the flexing of the outstanding legs. Erection bolts are placed near the bottom of these legs so as not to interfere with the bending.

Another type employs a T section whose flange is welded to the supporting member with two vertical welds. The stem overlaps the beam web and is attached by a vertical fillet weld along the edge, supplemented by a weld along one side of a vertical slot cut in the stem. Flexibility results from the bending of the T flange.

Seat connections. A bracket or shelf on which the end of the beam rests is a seat connection; it is intended to furnish the end reaction of the supported beam. The bracket may be attached to the support by rivets or welding. Two general types are used: the unstiffened seat provides bearing for the beam by a projecting plate or angle leg which offers resistance only by its own flexural strength; the stiffened seat is supported by a vertical plate or angle which transfers the reaction force to the supporting member without flexural distortion of the outstanding seat.

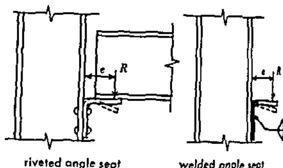


Fig. 3. Unstiffened seat connections.

Unstiffened seat connections. A simple form consists of an angle sufficiently long to engage the flange width, with its vertical leg attached to the support and the outstanding leg serving as the end bearing for the beam (Fig. 3). The unstiffened angle acting as a cantilever is suitable for small loads.

In riveted angle seats, the size of the angle must be such as to permit sufficient rivets in the vertical leg. The outstanding leg must project enough to distribute the beam reaction and thus avoid crippling of the beam web. Thick angles tend to concentrate the reaction near the outer edges as the beam rotates. The distribution is greater for thinner angles which bend with beam rotation (Fig 3). The uncertainty of the reaction eccentricity and the effect of connecting the beam flange to the seat require approximations in the analysis. To provide lateral stability, an angle is usually connected to the upper flange of the beam, permitting rotation without shear resistance.

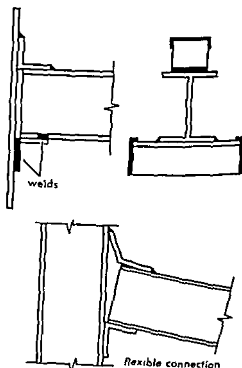


Fig. 4 Welded top angle and seat connection

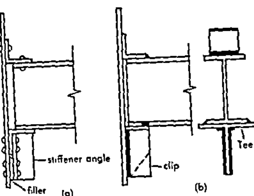


Fig 5 Stiffened seat connections. *(a)* Riveted stiffened seat *(b)* Welded stiffened seat.

In welded angle seats, the angle forming the seat is attached to the support by vertical fillet welds along the edges with a short return to the top. These welds are subjected to shear and bending forces due to eccentricity of load. The beam flange is attached to the seat by short edge welds (Fig. 4). The top angle is attached by fillet welds along the outer edges of both legs. Bending of the seat and flexing of the top angle provides a dependable amount of rotation of the beam.

Stiffened seat connections To provide sufficient rivets to resist large reactions and also to reinforce the seat against bending, one or two vertical angles, attached to the supporting member, are fitted tightly against the underside of the seat angle leg The legs of the stiffener angles extend to the outer edge of the seat The thickness of these angles provides contact area with the seat to limit the compression stress and avoid local buckling A filler plate is placed back of the extended stiffener angles (Fig 5a) The rivets connecting the stiffeners resist vertical and transverse shear. Tension is produced in the upper rivets and bearing pressures develop at the bottom of the stiffener owing to the moment of the eccentric load

Welded stiffened seats are made with a length of structural T, attached to the support by vertical fillet welds on both sides of the stem and additional welds along the underside of the flange to provide torsional stiffness (Fig 5b). A short weld attaches the beam [illegible]

[illegible]

concentrate the reaction force near the outer edge of the seat The vertical welds resist shear and moment

Eccentrically loaded connections. When the action line of a transferred force does not pass through the centroid of the connecting rivet group or welds the [illegible]
moment [illegible]
stresses in [illegible]
diagonal [illegible]
through a [illegible]
to the connecting rivet group, as a consequence, moment *Pe* tends to rotate the plate (Fig. 6a). Each rivet must resist its share of *P* plus the force induced by the moment, which is proportional to the distance from the centroid. These forces are combined vectorially to find the resultant rivet force (Fig 6b). Similar conditions exist in a column bracket or connections required to transmit moments caused by lateral forces on a building frame (Fig 6c). In welded connections the maximum stress in the weld is also the effect of the combined shear and moment.

Rivets in tension. In stiffened seat connections or when angles transfer load from a gusset plate to the face of a column, some rivets are subjected to tension induced by moments. No initial tension exists in cold-driven rivets, and the tension produced by applied moment is found by the flexure formula after locating the neutral axis.

Hot-driven rivets have initial tensile stresses in the shank as a result of restraint of contraction during cooling. Tests have shown there is a residual stress of about 24,000 ksi accompanied by compression on the contact surfaces of the connected parts Applied moment increases the rivet tension and decreases the accompanying compression. Analysis shows that the final tension is not appreciably greater than the initial tension and the connected parts do not separate under usual working loads.

When rivets are subjected to both shear and tension the resultant maximum stresses are computed by the combined stress formulae for maximum

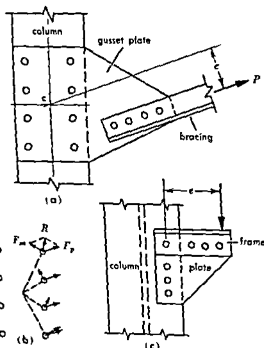


Fig. 6 Eccentrically loaded connections (a) Diagonal bracing (b) Resultant rivet forces (c) Lateral forces to column from frame produce eccentric loading

principal stress and maximum shear stress (*see* STRESS AND STRAIN). The maximum combined stress will be greater than the stresses caused by the tensile or shear forces acting alone. Some specifications limit the maximum combined stress to the value permitted in shear acting alone.

Moment-resisting connections. Rigidity and moment resistance are necessary at the ends of beams forming part of a continuous framework which must resist both lateral and vertical loads. Wind pressures tend to distort a building frame, producing bending in the beams and columns which must be suitably connected to transfer both moment and shear. The resisting moment can be furnished by various forms of angle, T, or bracket connections.

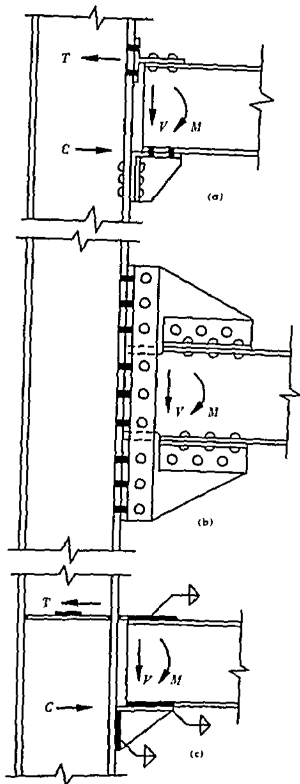


Fig 7. Moment resisting connections. (a) Seat and T or two Ts. (b) Bracket or stiffened beam seat. (c) Welded type.

A riveted connection consisting of web framing angles supplemented by angles connected to both top and bottom flanges is suitable for small end moments. The web angles are assumed to transfer the shear, while the flange angles connected to the column apply forces to the beam flanges which form the resisting couple. The end moment is limited by the number of rivets which can be provided in the legs of the flange angles and by the flexibility of the angles. A similar connection utilizes web angles and structural T sections for connection of both beam flanges, or a stiffened beam seat connecting the bottom flange with a T attached to the top flange (Fig. 7*a*). The tensile and compressive force required to furnish the resisting couple may be reduced by increasing the lever arm; the increase is accomplished by attaching short lengths of beams called beam stubs to one or both flanges, thus in effect deepening the connected beam and increasing the arm between the connecting Ts (Fig. 7*b*).

Welded moment-resisting connections are commonly made with a stiffened seat and a plate fillet welded to the top flange and butt-welded to the column flange. The column flanges are reinforced against bending by stiffener plates welded between the column flanges (Fig. 7*c*).

Pinned connections. Where appreciable angular change between members is expected, and in special cases where a hinge support without moment resistance is desired, connections are pinned. Many bridge trusses and large girder spans have pin supports. Plates connected to the adjacent members or the webs of the members themselves engage the pin in much the same manner as a bolted connection. As with riveted or bolted joints, the factors in design include bearing, shear, and bending of the pins, the net section of the plates or connected members, and edge tear out. Dishing or buckling of the plates must be avoided. The size of the pin is usually determined by its bending resistance. Reinforcing pin plates may be required to provide sufficient bearing.

High-strength bolts are made from heat-treated steel with a yield strength up to 90 ksi and ultimate strength of about 125 ksi. To develop high initial tension, the bolt is tightened with a calibrated torque wrench. Hardened washers under the head and nut distribute pressure to the assembled plates. The bolted joint transfers load by friction between the plates, developed by the high bolt tension. Because the holes are larger than the bolt diameter and the displacements are small, the bolts are not subjected to shear or bearing. Tests have shown that joints with high-strength bolts have a higher fatigue strength n stay tighter

longer than riveted joints under the same loading conditions. See BOLT; JOINT (MECHANICAL); RIVET; WELDING AND CUTTING OF METALS [W J KR.]

Bibliography: E. H. Gaylord, Jr. and C. N. Gaylord, *Design of Steel Structures*, 1957; L. Grover, *Manual of Design for Arc Welded Structures*, 1946; C. D. Williams and E. C. Harris, *Structural Design in Metals*, 2d ed., 1957.

Structural deflections

The deformations or movements of a structure and its flexural members, such as beams and trusses, from their original positions are called deflections. It is as important for the designer to determine deflections and strains, as it is to know the stresses that loads cause. See STRESS AND STRAIN.

Deflections may be computed by any of several methods. Generally the computation is based on the assumption that stress is proportional to strain. As a result, deflection equations involve the modulus of elasticity E which is a measure of the give of a material.

The relation between deflections at different parts of a structure is given by Maxwell's law of reciprocal deflections. This states that *if a load P is applied at any point A in any direction a* and causes a shift of another point B in direction b, the same load applied at B in direction b will cause an equal shift of A in direction a (*see illustration*). The law is used in a number of ways such as in simplifying deflection calculations, checking the accuracy of computations, producing influence lines. See STRUCTURAL ANALYSIS.

Beam and truss deflections usually are computed by similar methods, except that integration is used for beam equations and summation for trusses. Beam deflection equations involve bending moments and moments of inertia. Truss deflection equations are based on the stresses and cross sectional areas of chords and web members. For example, the dummy unit load equation takes this form for beams:

$$\delta = \int \frac{Mm\,dx}{EI}$$

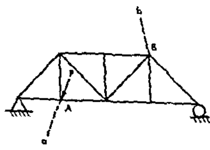
where δ is the deflection, M the bending moment due to applied loads, m the moment due to the dummy unit load applied at the point where the deflection is to be obtained, E the modulus of elasticity, and I the moment of inertia of a cross section of the beam. For trusses, it takes this form:

$$\delta = \sum \frac{SuL}{AE}$$

where δ is the deflection, S the stress in pounds in any member due to the actual load, u the stress in the member due to the dummy unit load applied where the deflection is to be obtained, L the length of the member, and A its area.

Deflections also may be determined graphically. The Williot-Mohr diagram, for example, often is used for trusses. It requires preliminary computation of stresses S and length changes SL/AE for truss members and plotting of these deformations to scale. [H.L.B.]

Bibliography: See STRUCTURAL ANALYSIS



Example of Maxwell's law of reciprocal deflections.

Structural geology

The branch of geology which, in the broadest sense, has to do with the description and analysis of the forms and interrelations of rock bodies. Structural features that came into being during the formation of the rocks are termed primary; those that were imposed upon the rocks after their formation are secondary. Although primary structures, such as bedding and igneous features, are considered in the study of structural geology, its major role is the description and analysis of secondary structures. *Tectonics* is synonymous with structural geology. Investigations of tectonic features involve their descriptions and interrelations, the analysis of the mechanics by which they were formed, and finally the synthesis of all structural data. See FAULT AND FAULT STRUCTURES; FOLD AND FOLD SYSTEMS.

A large body of information has accumulated describing the general forms of various secondary structures. However, most of these investigations have [illegible] displacements on fractures, or the internal distortion in rocks. Considerably more progress has been made in understanding the interrelationships of various structures, and undoubtedly this is the direction in which structural geology has made its greatest advance. A vast amount of information has been compiled on the relationships of minor to major structures, and certain of these relations have been sufficiently generalized to allow their use in inferring the presence of major tectonic features. See TECTONIC PATTERNS.

Structures have time as well as space relations, and the study of the sequential development of tectonic features generally proceeds hand in hand with the study of their spatial relations. Such studies attempt to ascertain the relative time of formation of various structures as well as to integrate the sequence of tectonic events into a unified geological history of the earth.

The study of the mechanics by which various secondary structures are formed has taken three general directions. One line of investigation has sought to determine the physical characteristics of rocks under various environmental conditions by appropriately designed experiments. A second approach has attempted to mimic, either by actual experiment or by theoretical analysis, the form of structures as they occur in nature And finally, attempts have been made to discover how the crystalline fabric of a rock deforms so as to accommodate itself to the general structural form. *See* ROCK MECHANICS; STRUCTURAL PETROLOGY.

Experimental tests. Of fundamental importance in understanding the deformation of rocks is the design of experiments to obtain information concerning the physical properties of rocks under various conditions. These data provide the basis for the mathematical theory of elasticity and plasticity in rocks. Toward this end compression tests, extension tests, and experiments involving the punching of disks have been performed under conditions of various pressures and temperatures on both minerals and rocks These experiments have furnished considerable information on the strengths of minerals and rocks and on the conditions under which they behave elastically or plastically. However, other factors are also important. Experiments in which the test block is immersed in a solvent yield different results from those obtained from tests in which the specimen is dry. The rate of application of the deforming pressure is also critical in controlling the nature of the deformation. Pressures and temperatures which are insufficient to cause plastic deformation in experiments in which the deforming pressure is applied rapidly are sufficient for plastic deformation when the same pressure is applied slowly. The information gained from these experiments is extremely important but as yet incomplete.

Experimental models. In mimetic investigations the correspondence of detail of form between the natural structure and the experimental or theoretical form serves to identify the mechanism of deformation In the simplest experiments the gross features of the natural structure are duplicated by a model in which the character of the material and the magnitude and distribution of pressures are controlled. Most of these experiments have limited significance because of the difficulty in scaling the factors involved in the natural structure down to the size of the model. An alternative approach utilizes certain assumptions concerning the stress distribution and the physical characteristics of rocks in the equations of elasticity and plasticity to compare the resulting theoretical form with that of the natural structure. This sort of device overcomes many of the problems of the experimental approach However, in practice, considerable difficulties are encountered. Much of the basic data concerning the physical characteristics of the rocks and the environmental conditions are inadequately known, and the equations of elasticity and plasticity are not sufficiently general to be applied to the complex structures in nature

Fabric analysis. Studies aimed at the elucidation of how the crystalline fabric accommodates itself to various structural forms have been based upon detailed statistical descriptions of the fabrics of natural rocks. Such investigations have demonstrated that minerals in deformed rocks may have shape orientation, internal structural orientation, or both. Additional information has been obtained from studies of fabrics imposed on rocks in the laboratory. These experiments have shown that intracrystalline deformation involving twinning and gliding has contributed to the over-all distortion of the rock, although without doubt intercrystalline deformation is also an important factor. In addition, observations of the fabrics of natural rocks suggest that the process of solution and recrystallization may be an important factor. Few experiments have been designed to test the effectiveness of this process, but attempts to approach the orientation of the internal structure of minerals from the thermodynamic standpoint have had limited success in certain simple cases, and this theoretical approach implies a process of solution and recrystallization. *See* PETROFABRIC ANALYSIS; PETROLOGY.

The ultimate goal of structural geology is the synthesis of all structural knowledge into a unified whole In the light of present knowledge, such a synthesis is somewhat unsatisfactory because it is necessarily based on incomplete and faulty data The geology of large tracts of the earth's surface is incompletely known, geophysical techniques provide but scanty information concerning the depths of the earth, and the mechanical basis for the formation of structures is only approximate. Nonetheless, attempts to compile the available data have value in that they help to bring the whole field of structural geology into focus and to stimulate investigations of critical problems. *See* ENGINEERING GEOLOGY; TECTONOPHYSICS. [P H O]

Bibliography: M P. Billings, *Structural Geology*, 2d ed , 1954; L. U DeSitter, *Structural Geology*, 1956, E. S. Hills, *Outlines of Structural Geology*, 3d ed., 1953; C. M. Nevin, *Principles of Structural Geology*, 4th ed , 1949.

Structural materials

Construction materials which, because of their ability to withstand external forces, are considered in the design of a structural framework. Materials used primarily for decoration, insulation, or other purposes are not included in this group

Structural clay products. The principal products in this class are the solid masonry units such as brick and the hollow masonry units such as clay tile or terra cotta.

Brick is the oldest of all artificial building materials It is classified as face brick, common brick,

and glazed brick. Face brick is used on the exterior of a wall and varies in color, texture, and mechanical perfection. Common brick consists of the kiln run of brick and is used principally as back-up masonry behind whatever facing material is employed. It provides the necessary wall thickness and additional structural strength. Glazed brick is employed largely for interiors where beauty, ease of cleaning, and sanitation are primary considerations.

Structural clay tiles are burned-clay masonry nits having interior hollow spaces called cells. uch tile is widely used because of its strength, ghtness in weight, and insulating and fire-protection qualities. Its size varies with the intended se.

Load bearing tile is used in walls that support, n addition to their own weight, loads that frame nto them, for instance, floors and the roof. Tiles manufactured for use as partition walls, for furring and for fireproofing steel beams and columns are classed as *nonload-bearing tile*. Special units are manufactured for floor construction. Some are used with reinforced-concrete joists, and others with steel beams in flat-arch and segmental-arch construction.

Architectural terra cotta is a burned-clay material used for decorative purposes. The shapes are molded either by hand in plaster of paris molds or by machine, using the stiff-mud process.

Building stones. The building stones generally used by the architect and engineer are limestone, sandstone, granite, and marble. Until the advent of steel and concrete, stone was the most important building material. Its principal use now is as a decorative material because of its beauty, dignity, and durability. See GRANITE; LIMESTONE; MARBLE; SANDSTONE; STONE AND STONE PRODUCTS.

Concrete. Concrete is a mixture of cement, mineral aggregate, and water, which, if combined in proper proportions, forms a fluid mass capable of being placed into molds and of hardening through the hydration of the cement. See CONCRETE.

Wood. The cellular structure of wood is largely what gives it basic characteristics unique among structural materials. The strength of wood depends on the thickness of the cell walls.

When cut into lumber, a tree provides a wide range of material which is classified according to use as yard lumber, factory and shop lumber, and structural lumber. Timber is lumber that is 5 [illegible] or larger [illegible]

Th [illegible] than [illegible] the ratio of its strength to its stiffness is much higher than that of steel or concrete; therefore, it is important that deflection be carefully considered in the design of a wooden floor system.

Laminated structural lumber is formed by gluing together two or more layers of wood with the grain of all layers *parallel* to the length of the member. Its principal advantages are the ease with which large members are fabricated and the greater strength of built-up members. Laminated lumber is used for beams, columns, arch ribs, chord members, and other structural members. See LUMBER MANUFACTURE; WOOD PHYSICS.

Structural metals. Of importance in this group are the structural steels, steel castings, aluminum alloys, magnesium alloys, cast iron, and wrought iron. See STRUCTURAL STEEL.

Steel castings are used for rocker bearings under the ends of large bridges. Shoes and bearing plates are usually cast in carbon steel, but rollers are often cast in stainless steel.

Aluminum alloys are strong, lightweight, and resistant to corrosion. The alloys most frequently used are comparable with the structural steels in strength. However, because aluminum alloys have a modulus of elasticity one-third that of steel, the danger of local buckling is likely to determine the design of aluminum compression members. Also, the accepted ratios of depth-to-span for bridges must be increased to reduce deflections and to give maximum economy of material. Because the alloy is approximately 35% the weight of steel, considerable savings in weight are achieved in long-span structures. Additional savings occur in the machinery, counterweights, and towers of bascule or lift bridges.

Magnesium alloys are produced as extruded shapes, rolled plate, and forgings. The principal structural applications are in aircraft, in the construction of truck bodies, and in the manufacture of portable scaffolding. Weight of the alloy is approximately 110 lb/ft^3.

Gray cast iron is used as a structural material for columns and column bases, bearing plates, stair treads, and railings. Malleable cast iron has few structural applications.

Wrought iron is used extensively because of its ability to resist corrosion. It is used for blast plates to protect bridges, for solid decks to support ballasted roadways, and for trash racks for dams. See CEMENT-ASBESTOS; GYPSUM PLANK. [C.N.G.]

Structural petrology

The study of rock fabric rather than of rock composition. In the usage of structural petrology, rock fabric includes (a) planar structural features, such as bedding, schistosity, cleavage, and joints; and (b) linear features, such as fold axes, axes of intersection of planar structures, parallel alignment of inequidimensional grains, stretching or smearing out of grains, and grooved slickensides. All these features can be observed in the field or on individual hand specimens without the aid of a microscope. Fabric also includes microfabric features recognizable under a microscope or by x-ray analysis, such as the size, shape, and arrangement of the constituent rock-making minerals and other components, as well as the lattice structure of rock-making minerals.

The study of the mechanics by which various secondary structures are formed has taken three general directions. One line of investigation has sought to determine the physical characteristics of rocks under various environmental conditions by appropriately designed experiments A second approach has attempted to mimic, either by actual experiment or by theoretical analysis, the form of structures as they occur in nature. And finally, attempts have been made to discover how the crystalline fabric of a rock deforms so as to accommodate itself to the general structural form. *See* ROCK MECHANICS; STRUCTURAL PETROLOGY.

Experimental tests. Of fundamental importance in understanding the deformation of rocks is the design of experiments to obtain information concerning the physical properties of rocks under various conditions. These data provide the basis for the mathematical theory of elasticity and plasticity in rocks. Toward this end compression tests, extension tests, and experiments involving the punching of disks have been performed under conditions of various pressures and temperatures on both minerals and rocks. These experiments have furnished considerable information on the strengths of minerals and rocks and on the conditions under which they behave elastically or plastically. However, other factors are also important. Experiments in which the test block is immersed in a solvent yield different results from those obtained from tests in which the specimen is dry. The rate of application of the deforming pressure is also critical in controlling the nature of the deformation. Pressures and temperatures which are insufficient to cause plastic deformation in experiments in which the deforming pressure is applied rapidly are sufficient for plastic deformation when the same pressure is applied slowly. The information gained from these experiments is extremely important but as yet incomplete.

Experimental models. In mimetic investigations the correspondence of detail of form between the natural structure and the experimental or theoretical form serves to identify the mechanism of deformation. In the simplest experiments the gross features of the natural structure are duplicated by a model in which the character of the material and the magnitude and distribution of pressures are controlled. Most of these experiments have limited significance because of the difficulty in scaling the factors involved in the natural structure down to the size of the model An alternative approach utilizes certain assumptions concerning the stress distribution and the physical characteristics of rocks in the equations of elasticity and plasticity to compare the resulting theoretical form with that of the natural structure. This sort of device overcomes many of the problems of the experimental approach. However, in practice, considerable difficulties are encountered. Much of the basic data concerning the physical characteristics of the rocks and the environmental conditions are inadequately known, and the equations of elasticity and plasticity are not sufficiently general to be applied to the complex structures in nature

Fabric analysis. Studies aimed at the elucidation of how the crystalline fabric accommodates itself to various structural forms have been based upon detailed statistical descriptions of the fabrics of natural rocks. Such investigations have demonstrated that minerals in deformed rocks may have shape orientation, internal structural orientation, or both. Additional information has been obtained from studies of fabrics imposed on rocks in the laboratory. These experiments have shown that intracrystalline deformation involving twinning and gliding has contributed to the over-all distortion of the rock, although without doubt intercrystal line deformation is also an important factor. In addition, observations of the fabrics of natural rocks suggest that the process of solution and recrystallization may be an important factor. Few experiments have been designed to test the effectiveness of this process, but attempts to approach the orientation of the internal structure of minerals from the thermodynamic standpoint have had limited success in certain simple cases, and this theoretical approach implies a process of solution and recrystallization. *See* PETROFABRIC ANALYSIS; PETROLOGY.

The ultimate goal of structural geology is the synthesis of all structural knowledge into a unified whole. In the light of present knowledge, such a synthesis is somewhat unsatisfactory because it is necessarily based on incomplete and faulty data The geology of large tracts of the earth's surface is incompletely known, geophysical techniques provide but scanty information concerning the depths of the earth, and the mechanical basis for the formation of structures is only approximate. Nonetheless, attempts to compile the available data have value in that they help to bring the whole field of structural geology into focus and to stimulate investigations of critical problems. *See* ENGINEERING GEOLOGY; TECTONOPHYSICS [P.H.O.]

Bibliography: M P Billings, *Structural Geology*, 2d ed , 1954, L U DeSitter, *Structural Geology*, 1956; E. S Hills, *Outlines of Structural Geology*, 3d ed , 1953, C. M. Nevin, *Principles of Structural Geology*, 4th ed , 1949.

Structural materials

Construction materials which, because of their ability to withstand external forces, are considered in the design of a structural framework. Materials used primarily for decoration, insulation, or other purposes are not included in this group.

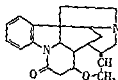
Structural clay products. The principal products in this class are the solid masonry units such as brick and the hollow masonry units such as clay tile or terra cotta.

Brick is the oldest of all artificial building materials. It is classified as face brick, common brick,

state only in portions of Africa. At one time it was widely raised in captivity for its ornamental plumage It is the largest living bird, and differs from all others in possessing a pubic symphysis and 2 toed feet The males weigh as much as 100 kg. It is highly adapted for running, having powerful legs and reduced wings. The breastbone is of the ratite or unkeeled type The head, neck, and thighs are almost featherless, and are brightly colored in breeding males As many as 20 eggs may be laid. Although they weigh about 1 3 kilograms each, they are small compared to the size of the parent. *See* AVES [K.C.P]

Strychnine

The principal alkaloid present in nux vomica, the seeds of a tree native to India, *Strychnos nux vomica* It was one of the first alkaloids to be isolated in a pure state in 1818 by P Pelletier and J. Caventou. The complex structure provided a fascinating problem which was pursued intensively for over a century, and was solved only in 1947. The synthesis

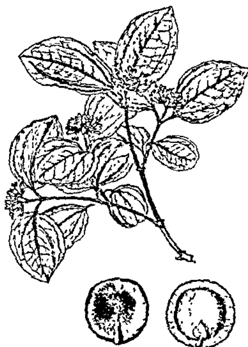


of strychnine by R Woodward and his coworkers in 1954 provided a confirmation of the structure.

Nux vomica was introduced into Germany in the sixteenth century as a poison for rats and other animal pests Strychnine was first employed in medicine in 1540, but it did not gain wide usage until 200 years later and has had an irregular career since then In the medical practice of an earlier day it had a reputation as a cardiovascular stimulant, respiratory stimulant, and bitter tonic Present-day opinion, however, holds that the therapeutically desirable effects are obtainable only with doses bordering on the toxic Pharmacological studies have shown that many of the therapeutic applications of strychnine have little or no rationale *See* ALKALOID [S M K]

Strychnos

A genus of tropical trees and shrubs belonging to the Logania family [illegible] *vomica* [illegible] of stry[illegible] dicinal [illegible] ders a[illegible] poison [illegible] and [illegible] from [illegible] islands Curare paralyzes the motor nerve endings in striated muscles and is used in medical practice in cases where a state of extreme muscular relaxation or even immobility is desirable It has become an important drug in the field of anaesthesiology. *See* GENTIANALES, STRYCHNINE. [F D.S]

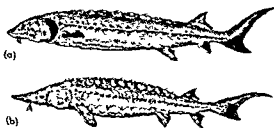


Strychnos nux-vomica Flowering branch and seeds. (From *H. W. Youngken, Textbook of Pharmacognosy,* 5th ed , Blakiston, 1946)

Sturgeon

Any of several similar primitive ganoid fishes with [illegible]

The largest sturgeon on record was caught in fresh water in Siberia and weighed 3000 lb; 2000-lb specimens of the great white sturgeon, *Acipenser transmontanus*, a marine species which spawns in fresh water, have been taken in the Columbia River. *A sturio*, the Atlantic sturgeon, occurs on the east coast of the United States. There are three fresh-water species in North America, all of which are becoming rare. Sturgeons live many years and mature slowly; *some* species do not produce eggs until they are 20 years old, or older



Sturgeon *(a)* Short-nosed, *Acipenser brevirostrum.* *(b)* American, *Acipenser sturio oxyrhynchus*, length to 12 ft (From *E L. Palmer, Fieldbook of Natural History,* McGrow-Hill, 1949)

By this study the relation of rock fabric to movements involved in the formation or deformation of the rock mass as a whole is established. *See* PETROFABRIC ANALYSIS. [E.B.K.]

Structural steel

Steel used in engineering structures, usually manufactured by either the open-hearth or the electric-furnace process. The exception is for carbon-steel plates and shapes whose thickness is 7/16 in. and under and which are used in structures subject to static loads only. These products may be made from acid-bessemer steel. The physical properties and chemical composition are governed by standard specifications of the American Society for Testing Materials (ASTM).

Structural carbon steel (ASTM-A7) is the best choice for ordinary riveted construction Since this steel is hardenable in the zone affected by the heat of welding it is not recommended for use in welded bridge construction. A373 steel, which has a base price somewhat greater than A7 steel, is preferred for the main members of all-welded bridges.

Structural nickel steel (ASTM-A8) contains 3–4% nickel and is not classed as weldable. It is used primarily for the main stress-carrying members of large structures.

Structural silicon steel (ASTM-A94) was first used in bridge construction in 1915. It has also been used for heavily loaded building columns. Because of its lower yield point and its hardness it is not as practicable as some of the other high-strength steels

Soft carbon-steel rivets (ASTM-A141) for riveting structural carbon-steel structures have been used since 1932 and are very satisfactory. Specification ASTM-A195 covers high-strength carbon-manganese rivets which, because of their hardness, are sometimes difficult to drive. ASTM-A406 high-strength rivet steel has a higher yield point than A195 and is free of the driving difficulties.

High-strength, low-alloy structural steel (ASTM-A242) is used where savings in weight and resistance to atmospheric corrosion are important. Fabrication is more difficult and a higher grade of workmanship is required than for A7 steel.

Manganese-vanadium steel and Tri-ten steel are recommended for welded bridge construction where high-strength and corrosion resistance are desirable.

T1 steel has the highest yield point (90,000 psi) of the structural steels. Its resistance to atmospheric corrosion is about four times that of carbon steel, and it has excellent welding properties. T1 steel can be used to advantage in the highly stressed members of large bridges. *See* STEEL; STRUCTURAL MATERIALS. [C.N.G]

Structures (engineering)

An engineering structure is a definite arrangement of related elements or members so connected as to support a given set of loads in a prescribed position. The principal structures designed by civil engineers are bridges, buildings, dams, docks, retaining walls, storage vessels, transmission towers, highway pavements, and aircraft landing strips. *See* AIRPORT ENGINEERING; BRIDGE; BUILDINGS; DAM; HIGHWAY ENGINEERING; PAVEMENT, RETAINING WALL.

A structure should be useful, safe, economical, and as attractive as possible. The design of a structure to fulfill these requirements is usually conceived in four stages which are seldom distinct but are carried along more or less simultaneously The first and frequently the most difficult phase is the development of the general layout to satisfy the functional requirements of the structure. Usually several solutions are prepared so all possible arrangements of the parts can be studied for the selection of the most satisfactory design. *See* CONSTRUCTION ENGINEERING.

The second major step in the design procedure is the development of the structural scheme. Since the functional plan may be greatly influenced by the need to eliminate potential structural difficulties as well as by the choice of materials and span lengths, the structural scheme is often developed during the functional planning stage Tentative cost estimates of several structural layouts will suggest the most economical scheme. Structural materials are selected not only on the basis of their structural properties but also on the availability of specific materials and skilled labor Relative costs and wage scales are also considered The character of a structure will often fix the choice of a material. Steel, aluminum, concrete, wood, and masonry have their own peculiar characteristics, and each is suitable for a particular form of construction *See* STRUCTURAL MATERIALS

The third stage of the design is structural analysis After the nature and magnitude of the loads to which the structure may be subjected are determined, the direct stress, shear, and moment in each member of the structure are calculated. This analysis requires a thorough understanding of the laws of statics, the theory of deflections, the principles of statically indeterminate structures, and the simplifying assumptions which must invariably be made *See* STRUCTURAL ANALYSIS

In the final phase of the design the members of the structure are proportioned to resist safely the internal forces disclosed by the structural analysis. In this stage the designer needs proficiency in predicting inelastic behavior under combined loadings in a design for repeated loading and in predicting buckling loads for inelastic eccentric columns He must also have a sound knowledge of strength of materials *See* STRENGTH OF MATERIALS [C.N.G]

Struthioniformes

An order of birds including but one family, Struthionidae, which contains the single living species of ostrich (*Struthio camelus*). Formerly occupying open areas of most of Africa, Arabia, and parts of western Asia, the ostrich is now found in a wild

roduction was the largest single end use for benne

The most common form of synthetic rubber, R S, is a copolymer of styrene with butadiene. roduction of polystyrene, used as a molding plasc, usually consumes at least half of the annual tyrene production

The polymerization of styrene is exothermic and hermally autocatalytic so that uncontrolled polymerization may occur with explosive violence Commercial styrene usually contains 4-(*tert* butyl)catehol to inhibit spontaneous polymerization

Styrene is a skin irritant Prolonged breathing of ir containing more than 400 parts per million of styrene vapor may be *injurious to health* *See* BENZENE, FRIEDEL-CRAFTS REACTION, POLYMERIZATION, POLYSTYRENE RESIN; RUBBER [C K B]

Bibliography: R H Boundy and R. F Boyer, (eds), *Styrene, its Polymers, Copolymers and Derivatives* 1952

Subgiant star

A member of the family of stars intermediate between giants and the main sequence in the Hertzsprung Russell (H-R) diagram (see STAR) The mean luminosity of a subgiant is about 10 times the Sun; the surface temperature lies between 7000 and 4000°K The masses are about 1 4 times that of the Sun The subgiants often violate the mass luminosity relation, that is, ζ Herculis A a G subgiant, is 4 times as bright as its mass would predict.

The subgiants are of particular importance in current theories of stellar evolution If a main-sequence star has exhausted about 12% of its mass of hydrogen the star begins to evolve, expanding cooling at the surface, and brightening Old stars of population II, of masses about 1 35 times the Sun, are now evolving into the subgiant region of the H R diagram The age of the oldest known stellar systems can be determined from the luminosities of the subgiants to be between 5 000 000,000 and 9 000 000 000 years. A different type of subgiant occurs also in close binary systems of the younger population I *See* STELLAR EVOLUTION [J L GR]

Subgraywacke

An argillaceous sandstone with a composition intermediate between graywacke and orthoquartzite (low rank graywacke of P D Krynine, lithic sandstones of F J Pettijohn) A clay matrix is usually present but in amounts less than 15% Unstable mineral and rock fragments are less than 25% Precise definitions of the boundaries of this group vary among sedimentary petrologists, but there is general agreement that these rocks contain moderate to large amounts of rock fragments, some clay matrix and at least a small amount of feldspar

The rock fragments in subgraywackes may be dominated by chert and other sedimentary species rather than metamorphic or igneous rocks The pore spaces are filled with a combination of clay matrix and mineral cement, usually quartz and carbonate The clay matrix is mainly muscovite (illite) with smaller amounts of kaolinite and, in some few cases, biotite and chlorite.

Subgraywackes are better sorted than the graywackes, partly because of the smaller amount of clay matrix and partly because the detrital sand-size fraction is well sorted. Detrital grains vary in roundness but tend to be rounded, in contrast to the angular grains of graywackes and the well-rounded grains of orthoquartzites Sedimentary structures are similar to those of the orthoquartzites but sometimes primary current lineation and groove and flute casts are found.

Subgraywackes are probably the most abundant sandstone type, and are found in deposits of all ages. They occur in moderately thick stratigraphic sections and are of wide lateral extent. They may be found both in geosynclinal and platform areas. The subgraywackes are found in association with micaceous and carbonaceous shales, thin biogenic limestones, and coal beds They seem to be characteristic of coastal plain and deltaic sedimentation and may be of mixed marine and nonmarine origin The source material, as is evidenced by the presence of some unstable minerals and sedimentary rock fragments, is a mixture of older sediments and perhaps some low-grade metamorphics This implies moderate tectonic activity in the source area but insufficient uplift to have allowed rigorous mechanical erosion to expose large areas of igneous and metamorphic rocks. *See* ARKOSE; GRAYWACKE; SANDSTONE; SEDIMENTARY ROCKS. [R S.]

Sublimation

The process by which solids are transformed directly to the vapor state without passing through the liquid phase Sublimation is of considerable importance in the purification of certain substances such as iodine, naphthalene, and sulfur.

Vapor pressure. All pure substances, whether solid or liquid, can exist in equilibrium with their vapor states, and the equilibrium pressure of the saturated vapor is called the vapor pressure of the solid or liquid at the temperature in question (*see* VAPOR PRESSURE) The area in the diagram to the right of the line *AOC* comprises the infinite number of pressure-temperature conditions for which a substance exists solely in the vapor state The area to the left of the line *AOB* represents the field of stability of the solid, in which the liquid or vapor cannot coexist with the solid Similarly, the area between the lines *OB* and *OC* is the field in which only the liquid phase is stable. The lines define the pressure-temperature conditions for the stable coexistence of pairs of phases. Thus, along the line *OA*, the solid and its vapor are in equilibrium, and the vapor pressure of the solid corresponding to any temperature along the line is unique The single intersection of the three vapor pressure curves at *O* is called the triple point and represents the only pressure and temperature at which the solid, liquid, and vapor can coexist in equilibrium under

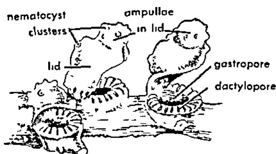
True caviar is made only from sturgeon roe, most of which comes from the Russian fishery in the Black Sea. Isinglass, now almost completely replaced by substitutes, is made from the swim bladder of the sturgeon. The flesh is highly prized because it is especially tasty when smoked. *See* ACIPENSERIFORMES; OSTEICHTHYES [J.D.B.]

Sty

An acute inflammation of one of the glands of the eyelids. A painful swelling appears at the root of an eyelash, usually an infection caused by a staphylococcus organism. The swelling results from pus formation, local edema, and obstruction of the gland duct The external glands of Moll and the internal meibomian glands may be affected A sty, or hordeolum, is usually associated with errors of refraction, poor general health, and other infections such as blepharitis, a generalized inflammation of the eyelids Drainage is usually spontaneous. In certain cases, the inner glands may become chronically obstructed and the inflammation will change to a small hard scar or nodule called a chalazion. *See* STAPHYLOCOCCUS. [E.C.ST.]

Stylasterina

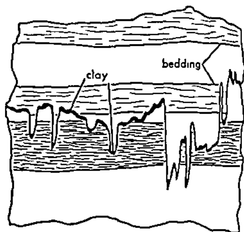
An order of the class Hydrozoa of the phylum Coelenterata, including several brightly colored branching or encrusting "corals" of warm seas. (True corals belong to a different class, the Anthozoa.) The calcareous skeleton is covered by living tissue and is penetrated by ramifying tubes Nutritive polyps, the gastrozooids, lie in cups on one surface or along certain edges of the skeletal substance. A spine, or style, at the base of each cup gives the order its name Associated with the gastrozooids are special stinging polyps, the dactylozooids, which have no mouths or tentacles Eggs and sperm are produced by special structures considered to be incompletely developed jellyfish. Some authorities combine the Stylasterina with the Milleporina in a single order, the Hydrocorallina. *See* HYDROZOA. [S CR.]



Cryptohelia. (From L. H. Hyman, The Invertebrates, vol. 1, McGraw-Hill, 1940)

Stylolites

Stylolites are irregular surfaces, mostly parallel to bedding planes, in which small toothlike projections on one side of the surface fit into cavities of like shape on the other side. A cross section of a typical stylolite would be similar to a profile of a series of high ridges and low valleys, with many of the peaks and valleys being about the same amplitude Stylolites are most common in limestones and dolomites but are also present in many other kinds of rock, including sandstones, gypsum beds, and cherts. Along almost all stylolite seams there is a thin layer of clay, quartz silt, or iron oxides, highly insoluble materials as compared with the rock proper. The amplitudes of the peaks and valleys range from a few millimeters to many centimeters.



Stylolite in limestone.

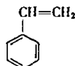
Stylolites cut across many structures in the rocks they traverse, including fossils and oolites Commonly these structures show truncation or partial solution along the stylolite seam The most probable origin for stylolites is pressure solution in an already lithified rock Large quantities of rock seem to have been dissolved along some stylolites and the thin insoluble coatings along the seams are apparently the insoluble residues. The insoluble residues are thicker at the peaks and valleys than in between, which also suggests a direct relationship to solution residues *See* DOLOMITE; LIMESTONE, SEDIMENTARY ROCKS. [R.S]

Bibliography: P B Stockdale, Stylolites: their nature and origin, *Indiana Univ Studies*, 11(55): 1–97, 1922.

Styrene

A colorless, liquid hydrocarbon which boils at 145 2°C and freezes at −30 6°C. It is also called vinylbenzene or phenylethylene. The ethylenic linkage of styrene readily undergoes addition reactions and under the influence of light, heat, or catalysts, will undergo self-addition or polymerization to yield polystyrene

$CH{=}CH_2$



Styrene is usually prepared industrially by the dehydrogenation of ethylbenzene obtained by alkylation of benzene with eth 56 styre

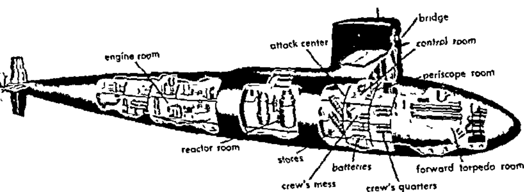


Fig 1 A nuclear-powered attack submarine (General Dynamics, Electric Boat Division)

leaves a wake. The nuclear reactor, which does not require air for its fuel, made the true submarine a reality As the submarine began to operate more extensively beneath the sea, its hull was modified in successive designs to give minimum resistance while running submerged These submerged-condition improvements were made at considerable sacrifice of performance on the surface. For a discussion of the factors which affect resistance, *see* SHIP PROPULSION

Classification Submarines are classified by their primary military missions Attack submarines (Fig 1) are fast, long-range vessels used primarily to detect and destroy merchant and naval ships Armed primarily with torpedoes, they may also carry mines which can be laid from the torpedo tubes.

Killer submarines carry more complex and sensitive underwater sound receivers (sonar) than attack submarines in order to fulfill their mission of detecting and destroying enemy submarines They are armed primarily with acoustic homing torpedoes and are generally smaller and slower than attack submarines

Radar picket submarines operate in exposed areas to give early warning of enemy aircraft and to direct air defense Their long-range, sky-searching radar and radio are more elaborate than those of other submarines

Guided missile submarines launch long range, airborne guided missiles from the surface, ballistic missile submarines launch long-range ballistic missiles either from the surface or while submerged

Midget submarines, armed with torpedoes or explosive charges make sneak attacks on ships in enemy harbors Because of their limited range and sea keeping ability, the midgets are either carried or towed by large submarines to the vicinity of their targets

Three types of submarine are designed to transport cargo or troops through enemy-controlled waters These are cargo, tanker, and troop transport submarines Tanker submarines were used in World War II to refuel attack submarines. With their limited carrying space and speed, as compared to surface ships, these three types of submarines are economical only when surface transportation is impossible or unacceptably hazardous.

Target submarines are used to train surface vessel crews and to develop new methods for detection and destruction of enemy submarines

Experimental submarines are built in limited numbers to test new features of hull shape, depth controls, or power plants for high-speed submarines.

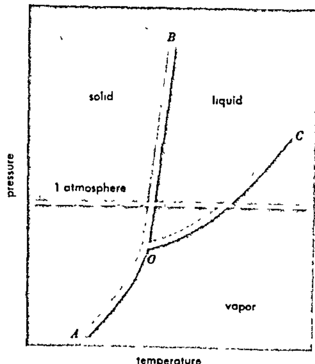
Equipment. Much of the equipment on the submarine, such as the electric plant, radio, sonar and radar equipment, and the hydraulic system, is similar in principle to that on the *surface ship.* The unique features of the submarine are discussed in the following paragraphs.

The outer hull is the external watertight boundary of the submarine A nonwatertight superstructure provides a smooth and fair envelope to cover the pipes, valves, and fittings on top of the hull Above the superstructure the fairwater similarly encloses the bridge, the periscope and mast supports, and, if one is provided, the conning tower. The inner hull is a second hull within all or a part of the length of the outer hull.

Pressure hull The pressure hull, comprising all of the inner and part of the outer hull, is the strong hull that resists external sea pressure when the submarine is submerged. Closed near the ends of the ship by flat or semiellipsoidal bulkheads, it is made up of cylinders and cones stiffened by internal bulkheads and frames. It is usually of circular or nearly circular cross-sectional shape for the best strength-to-weight ratio.

The pressure hull contains the ship's machinery and provides living quarters for the officers and crew. It is divided into watertight compartments that are further divided by platform decks into spaces for equipment and crew facilities. The main ballast tanks and fuel oil tanks are built into the area between the inner and outer hulls These tanks must be kept full of liquid so their internal pressure equals the sea pressure, preventing collapse of the relatively weak outer hull when the submarine is submerged.

Main ballast tanks. These tanks are so designated because they carry most of the water ballast;



Vapor-pressure–temperature diagram for a pure substance.

the pressure of the vapor alone. It is termed an invariant point, since neither temperature nor pressure may be varied without the disappearance of one of the phases The lines are univariant, since variation of either temperature or pressure independently does not cause the disappearance of a phase. A system of a pure substance in equilibrium between two phases is therefore said to possess one degree of freedom The areas (single-phase regions) are bivariant, and possess two degrees of freedom, since both temperature and pressure may be independently varied without the disappearance of the phase. *See* EQUILIBRIUM, PHASE.

Sublimation is a universal phenomenon exhibited by all solids at temperatures below their triple points. For example, it is a common experience to observe the disappearance of snow from the earth's surface even though the temperature is below the freezing point and liquid water is never present The rate of disappearance is low, of course, because the vapor pressure of ice is low below its triple point Sublimation is a scientifically and technically useful phenomenon, therefore, only when the vapor pressure of the solid phase is high enough for the rate of vaporization to be rapid. Necessarily, this is a relative consideration.

Triple point. For most substances, the triple point occurs at a comparatively low pressure, and the rate of sublimation is accordingly low For example, the triple point of water occurs at 0 0075°C and 4 56 mm pressure. For iodine, on the other hand, the triple point occurs at 114.15°C and 90 0 mm pressure. Accordingly, its rate of sublimation at 110°C would be quite high. In fact, the rate of sublimation is fast enough so that the iodine disappears by direct sublimation of the solid before the melting point is reached The liquid phase is observed only if the vapor is confined in a vessel. For a relatively few substances, the triple point lies above 1 atm pressure, and the vapor pressure of the solid attains atmospheric pressure before the liquid phase appears. Thus, dry ice (solid carbon dioxide) cannot be transformed to liquid CO_2 at atmospheric pressure. Instead, the solid sublimes to gaseous CO_2 without the intervention of the liquid state. The triple point of carbon dioxide is 5.11 atm at −56.4°C. The vapor pressure of solid CO_2 equals 1 atm at −78°C, on the other hand. Thus, the freezing point is higher than the sublimation point, and carbon dioxide does not possess a normal boiling point.

Energy requirements. Both the vaporization of a liquid and the sublimation of a solid require the absorption of heat to overcome the potential energy of the molecules in the condensed state. The molar latent heat of sublimation is completely analogous to the molar latent heat of vaporization. It is equal to the heat of vaporization of the liquid plus the heat of fusion of the solid. Moreover, the Clausius-Clapeyron equation describes the variation of the vapor pressure P of the solid with temperature T in similar fashion:

$$\frac{dP}{dT} = \frac{\Delta H_s}{T\,\Delta V}$$

where ΔH_s is the molar latent heat of sublimation, and ΔV is the difference in molar volume of the vapor and solid at the temperature T. If the vapor obeys the ideal gas law ($PV = RT$), this equation may be put into the form

$$\log_{10}\frac{P_2}{P_1} = \frac{\Delta H_s}{2.3R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)$$

which may be used to calculate the latent heat of sublimation if the vapor pressure is known at two temperatures, or the vapor pressure at a second temperature if the latent heat of sublimation and the vapor pressure are known at one temperature *See* EVAPORATION; MASS-TRANSFER OPERATION; SEPARATION (CHEMICAL AND PHYSICAL). [N H N]

Submarine

A ship that can operate both on the surface of the water and completely submerged Throughout most of World War II, submarines operated primarily on the surface and submerged only in the final stage of an attack or to evade detection. Since high surface speed was essential in this type of naval warfare, submarine hulls were designed for minimum surface resistance. The increasing effectiveness of radar and air patrols in locating and attacking surfaced submarines eventually forced them to remain submerged for longer periods In the late stages of the war the snorkel—a breathing tube which permits the submarine to get air for its diesel engines without surfacing—was developed However, even a snorkeling submarine may be detected visually, especially by aircraft, because it is submerged only slightly a [illegible] snorkel tube

pressure hull, is normally empty when the submarine is submerged. It may be flooded rapidly to give negative buoyancy for quick dives or to prevent the submarine from surfacing if depth control is lost.

Some main ballast tanks may be used to carry fuel oil. They are then called fuel ballast tanks and have connections to the fuel oil system and valves to close their flood holes

Periscopes. A periscope is an optical instrument in a tube 30–40 ft long with prisms and watertight windows at the upper and lower ends (see PERISCOPE) The main tube passes through a stuffing box at the top of the hull and may be housed within the fairwater or raised until the top of the tube is about 20 ft above the highest fixed part of the ship. Periscope depth is the greatest depth of the submarine at which the upper window of a fully raised periscope extends above the surface of the water. The upper prism may be tilted to permit use of the periscope while the submarine is rolling or pitching, or to observe aircraft. Auxiliary lenses within the tube may be moved in or out of the line of sight to make the magnification either 1 5 or 6 A telemeter scale in the field of view assists in estimating range, and an azimuth circle furnishes a scale for measuring bearings In a normal periscope, the eyepiece lowers with the tube when it is housed, and the periscope can thus be used only when extended to nearly its full height A more elaborate type has its eyepiece fixed in position and can be used throughout the travel One type of periscope, known as an attack periscope has the upper end of its tube tapered to a diameter of 1½ in or less to reduce the danger of detection Another type intended for use at night has an especially large upper window

Torpedo firing On all submarines the torpedoes are loaded into the torpedo tubes from the interior of the ship through breech doors After a tube is loaded and the breech door closed, a muzzle door is opened and the torpedo fired by admitting compressed air to the tube behind the torpedo The breech and muzzle doors are interlocked so that neither can be opened unless the other is closed Another interlock prevents firing unless the muzzle door is fully open

Escape system Every submarine has an escape system which allows the crew to leave the ship if it becomes disabled and cannot return to the surface The forward and after compartments are refuge compartments stocked with emergency food and water supplies, oxygen carbon dioxide absorbent, and inflatable life jackets Each refuge compartment of large submarines usually has an escape lock with an upper hatch set into the side of the lock and a lower access hatch leading from the refuge compartment The flooding of the lock and the opening of the upper hatch permits the crew to escape in small groups An air bubble remains in the lock in the space above the upper hatch After a group has departed the outer hatch may be closed from within the submarine and the lock drained and used again. On smaller submarines a single hatch, with a skirt extending down into the hull, is provided in each refuge compartment. Escape in this case is accomplished by flooding the entire compartment and opening the hatch The skirt below the hatch traps an air bubble in the top of the compartment so that each man has air while awaiting his turn to go through the hatch. With either method of egress, after the hatch is open each man in turn inflates his life jacket, takes several deep breaths, and goes through the hatch, continuing to exhale as the air in his lungs expands during the rapid ascent to the surface. This is called the buoyant ascent method.

Rescue chamber. If the submarine is unable to surface and weather conditions are favorable for rescue, a messenger buoy located at either end of the submarine can be released. The buoy carries one end of a downhaul cable to the surface, the other end being attached to the top hatch of the refuge compartment. Surrounding this hatch is a seat on which a rescue chamber will fit with a watertight joint Rescue chambers are carried by

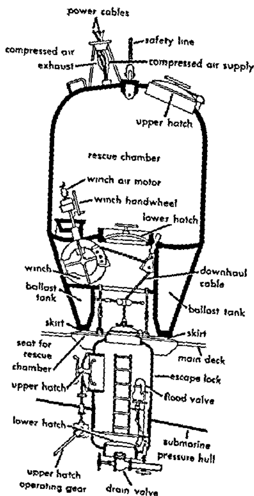


Fig 3 Submarine rescue system. (General Dynamics, Electric Boat Division)

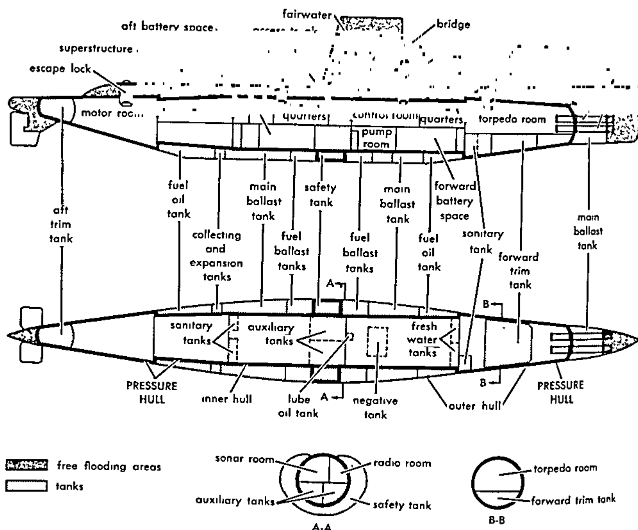


Fig. 2. Arrangement of compartments and tanks in a submarine (*General Dynamics, Electric Boat Division*)

they have large openings, called flood holes, at the bottom, and vent valves at the top The submarine is submerged by opening the vents, allowing air to escape and water to fill the tanks through the flood holes It is brought to the surface by closing the vents and admitting compressed air to the tanks to force enough water out through the flood holes to bring the superstructure awash The blowing is completed with air from a low-pressure air compressor or with diesel engine exhaust gases, thus conserving high-pressure air If an emergency occurs—for example, if the ship dives out of control—the main ballast tanks can be blown completely and rapidly with high-pressure air.

Diving planes. When, in the submerged condition, the weight and longitudinal position of the center of gravity are equal to those of water displaced by the submarine, the ship is said to be trimmed. As long as the submarine is moving it does not have to be perfectly trimmed; diving planes are used to balance moderate errors Diving planes are pairs of hydrofoils which extend from the sides of the ship, the bow planes at the forward part of the ship, the stern planes at the after end. On some ships the forward planes are mounted on the fairwater and are called fairwater planes. Each set is mounted on a horizontal stock and may be tilted through an angle of 25° in either direction from the horizontal to develop a vertical force on the planes and thus on the submarine. The angles that must be set on the diving planes to maintain constant depth give an indication of errors in the trim which are then corrected by changing the amount of water in the variable ballast tanks—the forward trim tank, the after trim tank, and the auxiliary tank See SHIP DESIGN

The combined capacity of the variable ballast tanks is such that, after the submarine has been brought to the desired over-all weight by permanent metal ballast, all changes in density of the sea water and in weight and longitudinal moments (due to the expenditure of fuel and other supplies) can be compensated for by varying the amount of water in these tanks. All variable ballast tanks are located within the pressure hull They have no direct connections to the sea, all pumping and flooding being done through a piping system known as the trimming line.

The safety tank, located within the pressure hull on its inboard side, is normally used as a main ballast tank With its flood and vent valves closed, it may be used as a variable ballast tank under an abnormal condition, for example, operation in fresh water. The negative tank, also inside the

motors as generators driven by the engine. The other propeller, driven by its engine, is normally used for propulsion at this time

On submarines with more engines than propeller shafts, some or all of the engines are connected only to generators. Any generator may be connected to the motors for propulsion or to the batteries for charging

Snorkeling The snorkel is a hollow mast or tube which may be extended above the fairwater to bring it above the surface of the water when the submarine is at periscope depth It provides air for combustion of fuel in the diesel engines At the top of the snorkel mast is a head valve, which shuts automatically when water rises above the air inlet and reopens when the water recedes. While the head valve is shut, air for the engines is drawn from the interior of the submarine. Snorkel safety devices stop the engines and close the exhaust valves if the pressure in the submarine falls below 12 pounds per square inch absolute The engines exhaust through a pipe that terminates a few feet below the top of the snorkel mast.

Nuclear propulsion While the nuclear powered submarine's steam propulsion machinery is gener-'ly similar to that of a surface ship, the necessity ·f operating this machinery at the high sea pres-ures of deep submergence requires special design ·r some components An additional difference is hat large air-conditioning units are needed to re-nove waste heat and vapor from the machinery paces, since access to the atmosphere is not avail-ble for their removal in the conventional manner, y ventilation Nuclear submarines are propelled y steam turbines that are connected through re-duction gears to the propeller shafts. An electric motor is also connected to each shaft for slow speed propulsion Each motor receives power from a storage battery, a turbogenerator, or a diesel-powered generator The steam turbine and the reduction gear are disconnected from the propeller shafts when the motors are in use

Steam for the main and auxiliary turbines is supplied from the reactor compartment, which may contain one or more reactors Water, pressurized to keep it from boiling, is pumped through the reactor and two closed loops, each containing a heat exchanger to which the water gives up the heat it obtained from the reactor. This water is radioactive and is called the coolant, its system is called the primary coolant system. Each heat exchanger, acting as a boiler generates on its secondary side saturated steam which passes through a steam drum and a water separator to the main steam line

The water in the steam system is completely isolated from the primary coolant to prevent the transfer of radioactive matter The reactor room may not be entered while the reactor is in operation. It is shielded to prevent the escape of harmful radiation so that the crew has complete freedom of movement throughout the rest of the submarine. After the reactor is shut down, the reactor room may be entered when radioactivity has decreased to a tolerable level. See REACTOR, SHIP PROPULSION; see also ANTISUBMARINE WARFARE; MARINE MACHINERY. [A.I.M.]

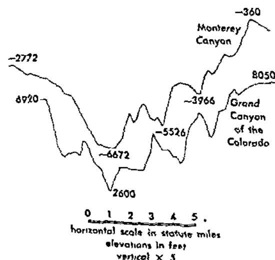
Bibliography: R. Blackman (ed.), *Jane's Fighting Ships, 1959–1960*

Submarine canyon

The sea floor has many puzzling features, but none which have aroused so much controversy as have the great submarine canyons, cut into the continental slopes off most coasts of the world Many of these have rocky walls thousands of feet high. They have narrow inner gorges, winding courses, numerous tributaries and are, in fact, quite comparable to the great canyons of the land (see illustration). Some of the canyons are direct continuations of land canyons, and others occur off large rivers which flow through broad flat-floored valleys on land. The submarine canyons extend outward down the slope virtually to the deep ocean floor. The outer portions of these sea valleys are of modest dimensions and extend across broad, gently sloping fans, comparable to the piedmont fans along the fronts of mountain ranges in arid regions. See CONTINENTAL SHELF AND SLOPE.

The deep floors of the canyons contain sediments alternating between sands, which resemble shallow water deposits, and normal deep-sea mud deposits It seems probable that landslides, occurring at the canyon heads, stir sediment into the water and produce a heavy suspension which sets up a current, called a turbidity current This moves along the canyon floors, leaving behind the sand deposits when the current loses velocity. These slides occur at rather frequent intervals, changing the depths and often breaking cables laid across the canyons. Thus cable companies avoid laying cables across submarine canyons where possible.

The cause of submarine canyons is much disputed. Their close resemblance to river-cut canyons



Transverse profile of the submarine canyon in Monterey Bay compared to a profile of the Grand Canyon of the Colorado River in Arizona.

submarine rescue vessels which, in peacetime, are stationed near submarine-operating areas.

The rescue chamber is a pressureproof cylinder with hatches in the top and bottom and a skirt extending about 3 ft below the lower hatch. It is ballasted to have a small amount of positive buoyancy and is hauled down to the submarine hatch by winding the downhaul cable on a winch inside the skirt. When the rescue chamber reaches the seat, the water is blown out of the skirt, a small tank is flooded to give the chamber negative buoyancy, and the pressure inside the skirt is equalized with the atmospheric pressure of the chamber. Sea pressure then seals the chamber to the seat and the submarine may be entered through the skirt and the submarine hatch. The submarine rescue vessel supplies power, light, communication, and compressed air to the rescue chamber through various cables and keeps a wire rope safety line attached to the chamber at all times. The chamber is manned by a crew of two and has room for nine others. After the chamber receives its passengers, an equal weight of ballast is discharged, the submarine hatch and the lower hatch of the chamber are closed, the small tank is blown to restore positive buoyancy, the skirt is flooded and equalized with sea pressure, and the rescue chamber returns to the surface by paying out the downhaul cable from its winch. Rescue is less dangerous than escape, especially in deep water, and is used when the equipment is available and weather conditions permit.

Propulsion. Most submarines in service today have two propellers, and use diesel engines for propulsion on the surface or at snorkel depth, and storage battery electric drive for underwater propulsion. With the ship submerged, each propeller shaft is turned through an engaged clutch by two dc electric motors in a single housing which receive power from two lead-acid storage batteries. In slow speed operations, the batteries are connected in parallel and the motors in series; for moderate speed both the batteries and the motors are connected in parallel; and for high speed the batteries are in series and the motors are in parallel. Speed variations within any of these combinations are made by varying the field strength of the motors. Under normal conditions, a diesel-electric submarine can remain completely submerged only as long as battery power is available. At slow speed, when the rate of battery discharge is low, this may be as much as three days, but at high speed the battery may be discharged in a few hours.

On some submarines the propellers may be driven by connecting the engines to the motors by mechanical clutches. The motors then either rotate idly or may be energized to supply a small charging current to keep the batteries fully charged (called "floating" the battery) while the submarine is in transit. Batteries are normally charged by disconnecting one of the propellers and energizing its

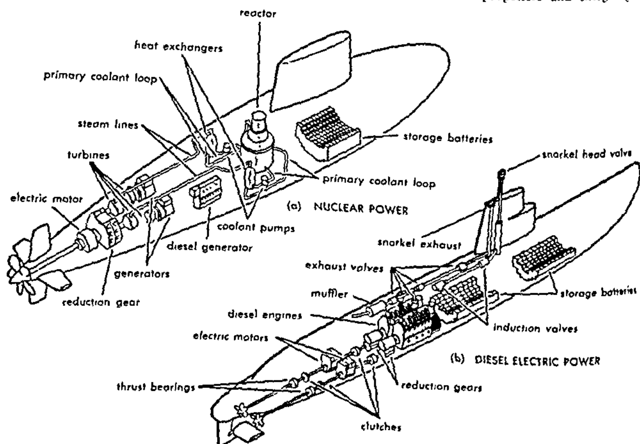


Fig 4 Submarine power plants. (a) Nuclear propulsion. (b) Diesel-electric propulsion. (General Dynamics, Electric Boat Division)

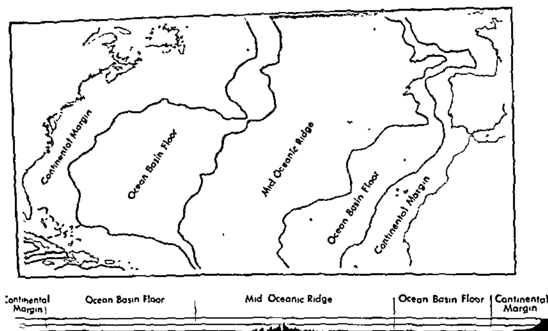


Fig 2. Major morphologic divisions of North Atlantic Ocean. The profile is from New England to Sahara.

- Earth's crust
  - Continents → Island arcs, Mediterraneans ← Oceans
  - Continents
    - Continental margin
      - I
        - Continental shelf
        - Epicontinental seas
        - Marginal plateaus
      - II
        - Shelf break
        - Marginal escarpment
        - Continental slope
        - Landward slope of trenches
          - Terraces
          - Scarps
      - III
        - Continental rise
          - Upper continental rise
          - Lower continental rise
          - Continental rise hills
          - Submarine canyons
        - Ridge-basin complex
        - Ridge trench complex
  - Oceans
    - Ocean basin floor
      - Abyssal floor
        - Abyssal hills
        - Abyssal plains
      - Oceanic rises
        - Scarp zones
        - Plateaus
      - Mid-ocean canyons
      - Abyssal gaps
    - Mid-ocean ridges
      - Flanks
        - Upper step
        - Middle step
        - Lower step
          - Scarps
          - Intermontane valleys
          - Nonseismic tilts
      - Crest
        - Rift valley
        - Rift mountains
        - High fractured plateau
    - Seamounts and islands
      - Archipelagic aprons
      - Flat tops
      - Terraces

Fig 3. Outline of submarine topography. Line 1, first-order features of the crust, line 2, major topographic features of the ocean, line 3, categories of provinces and superprovinces, line 4, provinces, line 5, subprovinces and other important features.

on land has convinced many geologists that they are due to river cutting followed by submergence of the valleys. The widespread distribution of submarine canyons has caused other geologists to object to this idea. Turbidity currents have been cited as an alternative cause. An unknown factor is the speed of turbidity currents, which may at times be very great although the evidence is not clear. If the canyons are caused by turbidity currents, it is difficult to understand why they should so closely resemble river-cut canyons. Furthermore, the existence of submarine canyons with hard rock walls, such as granite, has caused much dissatisfaction with the turbidity current hypothesis. Most geologists now agree that, however formed, submarine sliding of material and turbidity currents at least prevent the filling of the canyons of the sea floor. See TURBIDITY CURRENT; see also SUBMARINE TOPOGRAPHY. [F.P.S.]

Submarine topography

Over 70 per cent of the earth's surface is covered by marine waters. Of the oceanic area (361×10^6 km^2) approximately 300×10^6 km^2 is contributed by the deep-sea floor; the remaining 60×10^6 km^2 represents the submerged margins of the continents. The distribution of elevations on the earth is shown in Fig. 1.

Soundings. Soundings are measurements of ocean depth made from ships. Early soundings were made with a lead attached to a hemp line; about 1875, the hemp line was replaced by piano wire. Since about the middle of the 1920s, virtually all deep-sea soundings have been made by echo sounding. The echo-sounding machine sends out a sound pulse (10–20 kc) and then times the interval from the sound pulse to the returning echo. The early sounders required manual operation, but since about 1935 automatic recording sounders, which plot a graph of depth versus time or distance, have been used almost exclusively. Since 1953 precision, high-resolution echo sounders have been used in increasing numbers. See ECHO SOUNDER.

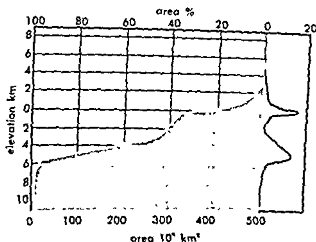


Fig. 1. Hypsographic curve showing area of earth's solid surface above any given level of elevation or depth. Curve at the right shows frequency distribution of elevations and depths for 2-km intervals.

Sounding corrections. Wire and hemp line soundings require a correction for wire angle, for stretch of the wire, and for calibration of the metering counters used. Echo soundings require a correction for sound velocity, since the average vertical velocity is not constant, and for slope of the bottom, as the point from which the first echo returns is not always directly beneath the ship. In addition, corrections for inaccuracies of timing and mechanical imperfections must be made for most sounders. The position of the sounding lines is generally determined by standard astronomical fixes and dead reckoning. Errors of a few miles are the rule in deep-sea sounding surveys. See CELESTIAL NAVIGATION; DEAD RECKONING; UNDERWATER SOUND.

PHYSIOGRAPHIC PROVINCES

The relief of the earth lies at two dominant levels (Fig. 1); one, within a few hundred meters of sea level, represents the normal surface of the continental blocks; the other, between 4000 and 5000 meters below sea level, and comprising over 50% of the earth's surface, represents the ocean basin floor. The topographic provinces beneath the sea can be included under three major morphologic divisions: continental margin; ocean-basin floor; and mid-oceanic ridge. These are indicated on a typical transoceanic profile taken from the North Atlantic in Fig. 2. Each of these major divisions can be further divided into categories of provinces and those into individual physiographic provinces (Fig 3).

Continental margins. The continental margin includes those provinces associated with the transition from continent to ocean floor. The continental margin in the Atlantic and Indian Oceans is generally composed of continental shelf, continental slope, and continental rise A typical profile off northeastern United States is shown in Fig 4 Gradients on the continental shelf average 1:1000, while on the continental slope gradients range from 1:40 to 1:6, and occasionally local slopes approach the vertical The continental rise lies at the base of the continental slope. Continental rise gradients average 1 300, but individual slope segments may be as low as 1:700, or as steep as 1:50. The continental slopes are cut by many submarine canyons Some of the larger canyons such as the Hudson extend across the continental rise (Fig 4). Submarine alluvial fans extend out from the seaward ends of the larger canyons. See CONTINENT; CONTINENTAL SHELF AND SLOPE; SUBMARINE CANYON

The continental margin can be divided into three categories of provinces Category I includes the continental shelf, marginal plateaus, and shallow epicontinental seas, all slightly submerged portions of the continental block Category II includes the continental slope, marginal escarpments, and the landward slopes of marginal trenches, all ex

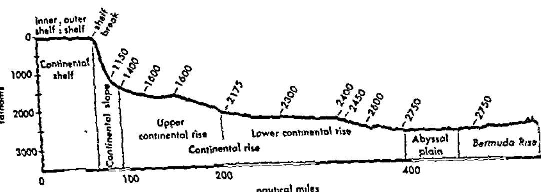


Fig 4. Continental margin provinces type profile off Northeastern United States.

buried beneath the abyssal plains. Abyssal plains are also found in the marginal trenches, marginal basins, and in epicontinental marginal seas Features of exactly the same morphology and origin are found in some lakes. Of similar origin are archipelagic aprons, which spread out from the base of oceanic islands *See* OCEANIC ISLANDS

Oceanic rises Oceanic rises are areas slightly elevated above the abyssal floor which do not be-

mic, their relief is more subdued and they are asymmetrical in cross section The western and central Bermuda rise is characterized by gentle, rolling relief The average depth gradually decreases towards the east. In the eastern third, the rise is cut by a series of scarps, 500–1000 fathoms in height, from which the sea floor drops to the level of the abyssal plain on the east. The series of eastward facing scarps suggest block faulting. Situated approximately in the center of the Bermuda Rise is the volcanic pedestal of Bermuda. A small archipelagic apron surrounds the pedestal. The turbidity-current origin of the smooth apron is supported by cores containing shallow-water carbonate clastic sediments in depths of 2300 fathoms. In the Pacific, extending for over 3000 miles west of Cape Mendocino, California, is an asymmetrical

cino Rise, but otherwise the relief of both features is quite similar. Although the circum-Pacific seismic belt crosses its trend, the Mendocino Escarpment is nonseismic. Other nonseismic fracture zones, which probably can be classified as oceanic rises, have been reported from the eastern Pacific. The Rio Grande Rise of the South Atlantic and the Mascarene Ridge of the Indian Ocean are similar in form.

Seamounts and seamount groups A seamount is any submerged peak over 500 fathoms high. This discussion, however, is limited to the larger, more or less conical peaks over 1000 fathoms in height. Seamounts are distributed through all the physiographic provinces of the oceans Seamounts sometimes occur randomly scattered, but more often lie in linear rows It seems safe to conclude that virtually all conical seamounts are extinct or active volcanoes. The Kelvin seamount group, a line of seamounts 800 miles long, stretches out from the vicinity of the Gulf of Maine toward the Mid-Atlantic Ridge. The Atlantis–Great Meteor seamount group extends for 400 miles along a north-south

a large area of seamounts whose flat summits range

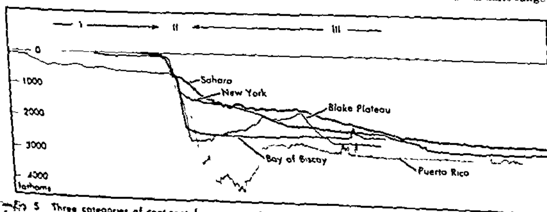


Fig 5 Three categories of continental-margin provinces.

pressions of the outer edge of the continental block. Category III includes the continental rise, the ridge-basin complex, and the ridge-trench complex. The continental slope of northeastern United States can be traced directly into the marginal escarpment (Blake Escarpment) off southeastern United States (Fig. 5), and the landward slope of the Antilles marginal trench (Puerto Rico). The continental rise off New England can be traced into the Antilles Outer Ridge. Seismic-refraction studies show that a trench filled with sediments and sedimentary rocks lies at the base of the continental slope off New England. Thus the main difference in morphology between the trenchless continental margins and those with a marginal trench is that in the former the trench has been filled with sediments.

In the continental margins of the Atlantic, Indian, Arctic, and Antarctic Oceans and the Mediterranean Sea, the continental rise generally represents the category III provinces. The Pacific, however, is bounded by an almost continuous line of marginal trenches. The high seismicity, vulcanism, and youthful relief of the Pacific borders suggest a very recent origin. In contrast, the nonseismic, nonvolcanic character, as well as the lower relief, of the Indo-Atlantic margins suggests a greater age. Thus on the old, stable, continental margins the deposition of sediment derived from the land has filled the marginal trench and produced the continental rise. The local relative relief on the continental margin rarely exceeds 20 fathoms, with the major exception of submarine canyons and occasional seamounts. See TECTONOPHYSICS.

Submerged benches. Submerged marine beach terraces have been identified throughout the world Since the beaches seem to correlate well between areas of vastly different tectonic development, it has been concluded that those listed in the accompanying table represent submerged Late Pleistocene beaches. See COASTAL LANDFORMS.

Structural benches. Structural benches, the topographic expression of outcropping beds, have been identified on the continental slope. Near Cape Hatteras, Virginia, the structural benches have been dated by extrapolating data obtained in several test borings near the coastline (Fig 6) Benches on Georges Bank have been dated from bottom samples obtained by dredging. Through the action of slumps, bottom currents, and turbidity currents, sediments are continually removed from the continental slope. Thus, there is no cover of recent sediments to obscure the outcrops of the ancient formations. See TURBIDITY CURRENT.

Ocean basin floor. Excluding the marginal trenches and mid-oceanic ridges, the deepest portions of the ocean are included in this division. Approximately one-third of the Atlantic and three-fourths of the Pacific fall under this heading The ocean-basin floor can be divided into three categories of provinces: the abyssal floor; oceanic rises; and seamounts and seamount groups.

Abyssal floor. The abyssal floor includes the broad, deep areas of the central portion of the ocean. In the Atlantic, Indian, and northeast Pacific Oceans, abyssal plains occupy a large part of the abyssal floor. An abyssal plain is a smooth portion of the deep-sea floor where the gradient of the bottom does not exceed 1:1000. Abyssal plains adjoin all continental rises and can be distinguished from the continental rise by a distinct change in bottom gradient. At their seaward edge, most of the abyssal plains gradually give way to abyssal hills. Individual abyssal hills are 50–200 fathoms high and 2–6 miles wide. In the Atlantic, the abyssal hill provinces only locally exceed 50 miles in width. Abyssal plains in the same area range from 100–200 miles in width. Core samples of sediment obtained from the Atlantic abyssal plain invariably contain beds of sand, silt, and gray clay intercalated in the red or gray pelagic clay which is generally characteristic of the deep-oceanic environment These deep-sea sands were transported by turbidity currents from the continental margin Some of the currents probably descended along a broad front, while others certainly followed the submarine canyons and spread out fanwise from their submarine alluvial cones. See MARINE SEDIMENTS

The abyssal hills are thought to represent tectonic or volcanic relief of a type identical with that

Depth in fathoms of prominent continental-shelf terraces*

Placentia Bay, Newfoundland	Norfolk, Virginia	Charleston, South Carolina	Bimini, B.W.I.	St Vincent, Cape Verde Is	Dakar, Senegal	San Pedro, California
10		12	10	8	10	10
				15	15	15
20	18	20	20	24	20	20
	30	30	28	28	28	28
35	35	35		32		
40				38	38	38
42		45	42	42	45	45
	50					
55	58			54	55	55
				60		
68		68	65			
80	80	80	85	80	78	80

* Each column based on a single nonprecision echogram.

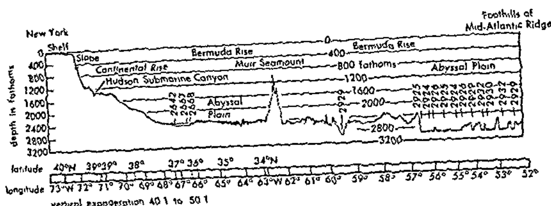


Fig 7. Precision-depth-recorder profile between Mid-Atlantic Ridge and New York, showing abyssal plains, Bermuda Rise, and the continental margin.

circle, if made with an average pencil point, will include all of the crust of the earth The crust has a density of about 2 7 The oceanic crust is about 6 km thick compared to about 36 km for the continents The boundary between the crust and the mantle is called the M discontinuity (Mohorovičić discontinuity) by seismologists

Unlike the continental areas the ocean floors cannot be studied directly, hence most of the information about the structure of the earth beneath the oceans comes from geophysical measurements Principally employed are earthquake seismology explosion seismology, and measurements of the variations in the earth's gravitational and magnetic ··ls Seismology is the study of the propagation ound or elastic waves in the earth. By measur- the speed of sound waves traveling in the vari- layers certain physical properties such as den- and elastic constants can be estimated. These icate the type of rock constituting each layer. travel time of sound waves reflected or re- cted by a particular layer provides a measure of depth to the layer Gravity measurements show nsity variations and are used with seismic evi- nce to indicate compositional changes or struc- ral features such as folds and faults Magnetic easurements give some evidence about the min- al constitution of rocks and are particularly use- l for structural mapping where volcanic and intrusive igneous rocks are present. The structure sections shown in Figs 9, 10, and 11 are based on information obtained by these techniques.

Figure 9 is a typical structure section from continent to ocean. This shows the relative thickness of the earth's crust (that portion above the M discontinuity) in the different areas On the continent and continental shelf the rocks beneath the sedimentary layers are granitic or granodioritic. Beneath is an intermediate layer believed to be gabbroic or basaltic. The mantle is probably peridotite a rock ...

is

frc

the center. There are some variations in the upper parts of the mantle between continental and oceanic areas, but the most apparent difference is in the thickness and composition of the crust. The continental crust is six or seven times thicker than the oceanic crust and contains almost all of the acidic rocks, such as granites, whereas the oceanic crust is almost entirely composed of basic rock.

Ocean basins. The average depth of the ocean basins is about 4 8 km The topography of the ocean floor is rough in the majority of the explored parts, although there are broad areas, particularly in the Atlantic, where the bottom is almost completely flat. These abyssal plains are thought to have been

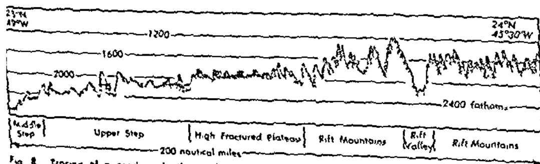


Fig 8 Tracing of a precision-depth-recorder record showing crest and western flank of the Mid Atlantic Ridge

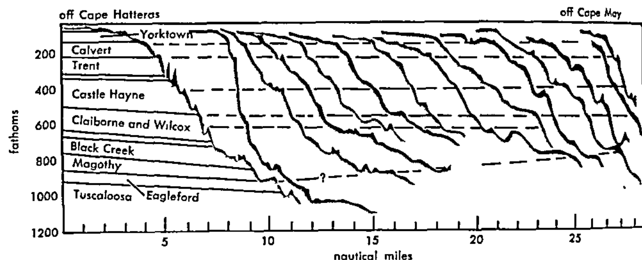


Fig. 6. Correlation of structural benches on the continental slope, Cape Hatteras to Cape May. Soundings by U.S Coast and Geodetic Survey.

from 50–850 fathoms beneath sea level. These tablemounts have been termed guyots From the flat summits, shallow-water fossils of Cretaceous age have been dredged Such sunken islands are not limited to the Pacific. Several of the Kelvin seamounts are flat-topped at 650 fathoms, and the seamounts of the Atlantis–Great Meteor group have flat summits at 150–250 fathoms *See* SEAMOUNT AND GUYOT.

Mid-oceanic ridge. The middle third of the Atlantic, Indian, and South Pacific Oceans is occupied by a broad, fractured swell known as the Mid-Oceanic Ridge In the Atlantic, it is known as the Mid-Atlantic Ridge, in the southern Indian Ocean it is the Mid-Indian Ridge, in the Arabian Sea it is the Carlsberg Ridge and Murray Ridge, and in the South Pacific it is known as the Easter Island Ridge.

The Mid-Atlantic Ridge can be divided into distinctive physiographic provinces which can be identified on most trans-Atlantic profiles (Figs 2 and 8).

Crest provinces. The rift valley, rift mountains, and high fractured plateau which constitute this category form a strip 50–200 miles wide The rift valley is bounded by the inward-facing scarps of the rift mountains. The floor of the rift valley lies 500–1500 fathoms below the adjacent peaks of the rift mountains which drop abruptly to the high fractured plateau, lying at depths of 1600–1800 fathoms on either side of the rift mountains The topography of the crest provinces is the most rugged submarine relief. An earthquake belt accurately follows the rift valley through a distance of over 40,000 miles. Heat-flow measurements in the crest provinces give values several times greater than have been obtained in normal ocean or continental areas. A large positive magnetic anomaly and a moderate (−20 milligals) negative gravity anomaly are associated with the rift valley. Seismic-refraction measurements indicate a crust intermediate in composition between the oceanic crust and the mantle. The crest provinces of the Mid-Oceanic Ridge can be traced directly into the rift valleys, rift mountains, and high plateaus of Africa These features of African geology are clearly the result of extensional forces in the earth's crust. The Mid-Oceanic Ridge is probably similar in all essential characteristics, including origin, to the African rift valley complex. *See* RIFT VALLEY

Flank provinces. The flank provinces of the Mid Oceanic Ridge can be divided into several steps or ramps, each bounded by scarps somewhat larger than those which characterize the entire area

Parts of the flank provinces, particularly the Upper Step south of the Azores, are characterized by smooth-floored intermontane valleys. Photographs, cores, and dredging indicate that the crest of the Mid-Atlantic Ridge north of the Azores is being denuded of its sediments. As these sediments are eroded from the crest provinces and deposited on either side, they are gradually filling the intermontane basins and smoothing the relief of the flanks. [B C H]

UNDERLYING STRUCTURE

Because approximately 70 per cent of the surface of the earth is covered by the oceans, the typical structure of the earth is found in the oceanic and not in the land areas Statistical examination shows that most of the earth's solid surface is either at the elevation of the ocean floors or at the elevation of the continents (Fig. 1). The anomalous areas, those of extreme or of intermediate elevation, are long, narrow features—the mountain ranges, island arcs, deep-sea trenches, and continental margins

To have a rough model of a section through the earth, one draws a circle about 5 in in diameter and a concentric one of about half that diameter Inside the smaller circle is the core, of very dense material, possibly metallic. The part between the circles is the mantle, a crystalline, basic rock with density about 3 3 g/cm³. The line forming the outer

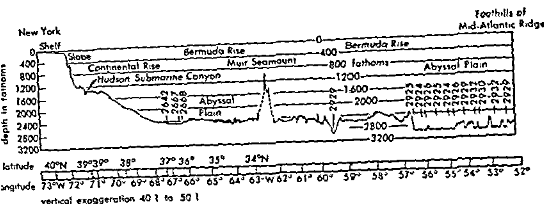


Fig. 7 Precision-depth-recorder profile between Mid-Atlantic Ridge and New York, showing abyssal plains, Bermuda Rise, and the continental margin

circle, if made with an average pencil point, will include all of the crust of the earth. The crust has a density of about 2.7. The oceanic crust is about 6 km thick compared to about 36 km for the continents. The boundary between the crust and the mantle is called the M discontinuity (Mohorovičić discontinuity) by seismologists.

Unlike the continental areas, the ocean floors cannot be studied directly; hence most of the information about the structure of the earth beneath the oceans comes from geophysical measurements. Principally employed are earthquake seismology, explosion seismology, and measurements of the variations in the earth's gravitational and magnetic fields. Seismology is the study of the propagation of sound or elastic waves in the earth. By measuring the speed of sound waves traveling in the various layers, certain physical properties such as density and elastic constants can be estimated. These indicate the type of rock constituting each layer. The travel time of sound waves reflected or refracted by a particular layer provides a measure of the depth to the layer. Gravity measurements show density variations and are used with seismic evidence to indicate compositional changes or structural features such as folds and faults. Magnetic measurements give some evidence about the mineral constitution of rocks and are particularly useful for structural mapping where volcanic and intrusive igneous rocks are present. The structure sections shown in Figs. 9, 10, and 11 are based on information obtained by these techniques.

Figure 9 is a typical structure section from continent to ocean. This shows the relative thickness of the earth's crust (that portion above the M discontinuity) in the different areas. On the continent and continental shelf the rocks beneath the sedimentary layers are granitic or granodioritic. Beneath is an intermediate layer believed to be gabbroic or basaltic. The mantle is probably peridotite, a rock principally composed of olivine. This is the most prominent layer in the earth, extending from near the surface approximately halfway to the center. There are some variations in the upper parts of the mantle between continental and oceanic areas, but the most apparent difference is in the thickness and composition of the crust. The continental crust is six or seven times thicker than the oceanic crust and contains almost all of the acidic rocks, such as granites, whereas the oceanic crust is almost entirely composed of basic rock.

Ocean basins. The average depth of the ocean basins is about 4.8 km. The topography of the ocean floor is rough in the majority of the explored parts, although there are broad areas, particularly in the Atlantic, where the bottom is almost completely flat. These abyssal plains are thought to have been

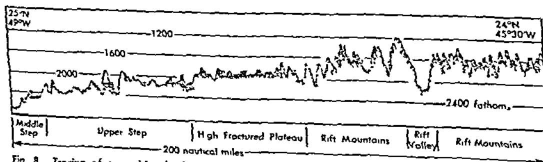


Fig. 8 Tracing of a precision-depth-recorder record showing crest and western flank of the Mid-Atlantic Ridge

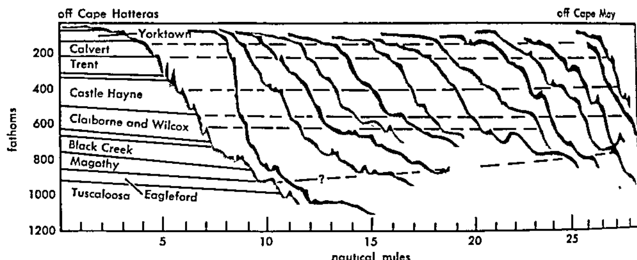


Fig. 6. Correlation of structural benches on the continental slope, Cape Hatteras to Cape May Soundings by U.S. Coast and Geodetic Survey.

from 50–850 fathoms beneath sea level. These tablemounts have been termed guyots From the flat summits, shallow-water fossils of Cretaceous age have been dredged. Such sunken islands are not limited to the Pacific. Several of the Kelvin seamounts are flat-topped at 650 fathoms, and the seamounts of the Atlantis–Great Meteor group have flat summits at 150–250 fathoms. *See* SEAMOUNT AND GUYOT

Mid-oceanic ridge. The middle third of the Atlantic, Indian, and South Pacific Oceans is occupied by a broad, fractured swell known as the Mid-Oceanic Ridge. In the Atlantic, it is known as the Mid-Atlantic Ridge, in the southern Indian Ocean, it is the Mid-Indian Ridge, in the Arabian Sea it is the Carlsberg Ridge and Murray Ridge, and in the South Pacific it is known as the Easter Island Ridge.

The Mid-Atlantic Ridge can be divided into distinctive physiographic provinces which can be identified on most trans-Atlantic profiles (Figs. 2 and 8).

Crest provinces. The rift valley, rift mountains, and high fractured plateau which constitute this category form a strip 50–200 miles wide. The rift valley is bounded by the inward-facing scarps of the rift mountains. The floor of the rift valley lies 500–1500 fathoms below the adjacent peaks of the rift mountains which drop abruptly to the high fractured plateau, lying at depths of 1600–1800 fathoms on either side of the rift mountains. The topography of the crest provinces is the most rugged submarine relief An earthquake belt accurately follows the rift valley through a distance of over 40,000 miles. Heat-flow measurements in the crest provinces give values several times greater than have been obtained in normal ocean or continental areas. A large positive magnetic anomaly and a moderate (−20 milligals) negative gravity anomaly are associated with the rift valley. Seismic-refraction measurements indicate a crust intermediate in composition between the oceanic crust and the mantle. The crest provinces of the Mid-Oceanic Ridge can be traced directly into the rift valleys, rift mountains, and high plateaus of Africa. These features of African geology are clearly the result of extensional forces in the earth's crust. The Mid-Oceanic Ridge is probably similar in all essential characteristics, including origin, to the African rift valley complex *See* RIFT VALLEY.

Flank provinces The flank provinces of the Mid Oceanic Ridge can be divided into several steps or ramps, each bounded by scarps somewhat larger than those which characterize the entire area

Parts of the flank provinces, particularly the Upper Step south of the Azores, are characterized by smooth-floored intermontane valleys. Photographs, cores, and dredging indicate that the crest of the Mid-Atlantic Ridge north of the Azores is being denuded of its sediments. As these sediments are eroded from the crest provinces and deposited on either side, they are gradually filling the intermontane basins and smoothing the relief of the flanks [B C H]

UNDERLYING STRUCTURE

Because approximately 70 per cent of the surface of the earth is covered by the oceans, the typical structure of the earth is found in the oceanic and not in the land areas Statistical examination shows that most of the earth's solid surface is either at the elevation of the ocean floors or at the elevation of the continents (Fig 1) The anomalous areas those of extreme or of intermediate elevation, are long, narrow features—the mountain ranges, island arcs, deep-sea trenches, and continental margins

To have a rough model of a section through the earth, one draws a circle about 5 in in diameter and a concentric one of about half that diameter Inside the smaller circle is the core, of very dense material, possibly metallic The part between the circles is the mantle, a crystalline, basic rock with density about 3 3 g/cm³. The line forming the outer



g 11 Structure section across Puerto Rico and the Puerto Rico Trench. Vertical exaggeration X10.

Trench, the Tonga Trench, the South Sandwich Trench, and the trenches of Japan and the East Indies. The greatest depths in the oceans are found in these trenches, the deepest in the Pacific being about 10.7 km in the Mariana Trench and in the Japan Trench, and the deepest in the Atlantic about 8.4 km in the Puerto Rico Trench. Several of these trenches have been investigated, using geophysical techniques, and found to be similar in most respects. Figure 11 is a structure section from the Atlantic Ocean into the Caribbean Sea crossing the Puerto Rico Trench. The trench is formed by the depression or the sinking of the high-velocity crustal layer and the mantle by several kilometers. In the bottom are layers of lower velocity, less dense materials which are probably sediments and volcanic debris from the nearby islands and perhaps sediments which have flowed in from the ocean basin. The great depth of the dense layers causes a pronounced deficiency of gravity, a characteristic feature of all deep-sea trenches.

Several hypotheses have been advanced to account for the existence of island arcs and the associated trenches and to relate them to some major process going on in the earth. For example, there is evidence that the continents have grown during geologic time, and many believe that they have been encroaching on the oceans through a process in which deep-sea trenches are formed near the continental margins and filled with sediments which are then metamorphosed, folded, and up-[illegible]ted to become mountain ranges. Similar reasoning could apply to island arc trenches. Speculation about the formation of trenches has followed [illegible]veral lines [illegible] impressi[illegible] on fault[illegible] convect[illegible] suggested [illegible] [illegible]ERIOR; GEODESY; GEOMAGNETISM; OROGENY; SEISMOLOGY; TECTONIC PATTERNS; TERRESTRIAL GRAVITATION [W.M.E.; J.I.E.]

OCEAN BASIN FORMATION

The earth's crust [illegible] topograph[illegible] represent [illegible] and cont[illegible] Geologists have long speculated concerning the permanence of this distribution and the possibility of frequent interchanges between deep sea and continent. The strong contrast between continental and oceanic crustal structure makes the assumption of frequent interchange seem improbable. However, paleontological evidence indicates former land connections between now widely separated continental masses. It has been suggested that long, narrow, land bridges were built across the oceans, which have since broken up and subsided into the ocean depths. See DEVONIAN; PALEONTOLOGY.

Another school of thought discounts the necessity of the supposed connections and maintains that each continent has been built up by a gradual accumulation of light crust around an original small nucleus. Sediments eroded from primordial shields are supposed to have filled marginal geosynclines which were later accreted to the protocontinental masses which in this way grew to their present size.

Some workers link the origin and distribution of continents and oceans to convective cells in the earth's mantle. Advocates of convection can be generally divided into two groups, one holding that currents rise under the continents, the other advocating rising currents under the ocean. The latter group maintains that these currents sweep the lighter products of differentiation toward the continents.

Alfred Wegener proposed that the present continents were originally part of a single mass, which broke up in the Cretaceous and whose parts drifted apart to their present positions. Recent work on paleomagnetism has given impressive support to this concept. However, students of submarine topography object to such a simple concept which ignores the topography of the deep-sea floor. Recently attention has turned to a theory not previously considered seriously. This theory proposes that the earth's center has expanded considerably during geologic time. This expansion resulted in the displacement of the fragments of the original differentiated crust to form the present continents. The theory seems to explain the continental displacement evidence of paleomagnetism and paleoclimatology, and in addition provides a mechanism for the formation of suboceanic relief. See PALEOCLIMATOLOGY; ROCK MAGNETISM. [B.C.H.]

Bibliography: A. L. DuToit, *Our Wandering Continents*, 1937; J. Ewing and M. Ewing, Seismic-refraction measurements in the Atlantic Ocean Basins, in the Mediterranean Sea, on the Mid-Atlantic Ridge, and in the Norwegian Sea, *Geol Soc Am Bull.*, 70:291–318, 1959; B. C. Heezen, M. Tharp, and M. Ewing, *The Floors of the Oceans: I, The North Atlantic*, Geol. Soc. Am. Spec. Paper 65, 1959; M. N. Hill, Recent geophysical explora[illegible] et [illegible] vol [illegible] H. [illegible] Northeastern Pacific Basin and the West Coast of North America, *Geol. Soc. Am. Bull.*, 66:1149–1198, 1955; F. P. Shepard and K. O. Emery, *Submarine Topography off the California Coast: Canyons and*

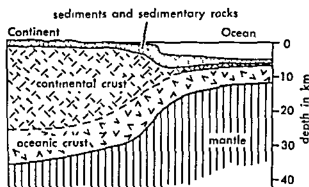


Fig 9. Structure section from continent to ocean. Vertical exaggeration X10.

formed by turbidity current deposition where sediments, set in motion during underwater landslides and thrown into suspension in the water, flow to the deepest parts of the basins. The average thickness of sediments varies from tens of meters on the elevations to thousands in the deeper parts. The average is about 1 km.

The layer immediately below the sediments (see Fig. 9) is difficult to identify. Only its seismic velocity is known, and that varies between 4 and 6 km/sec. This range of velocities encompasses those appropriate to compacted or metamorphosed sediments, volcanic rocks, or continental granitic rocks. The layer could be composed of any or all of these materials, or it might be thought of as scum on the top of the underlying main crustal layer.

The principal layer of the oceanic crust, shown by check symbols in the structure sections, has been found by all of the numerous seismic-refraction measurements made in the Atlantic, Pacific, and Indian Oceans. The velocity of sound in this layer is consistently near 6.6 km/sec and the thickness is about 5 km. Variations from these averages are sometimes found in the neighborhood of anomalous areas such as seamounts, trenches, or continental margins. A widely held belief is that this layer is the primordial outer surface of the earth Whether or not it continues under the continents is under dispute. There is much evidence that the velocity in the lower part of the continental crust is intermediate between that in the upper part and that in the mantle. The intermediate material may represent the continuation of the main oceanic crustal layer. In view of its seismic velocity and of the fact that rocks brought up in oceanic volcanoes are predominantly olivine basalts, the main crustal layer is commonly called the basalt layer. A study is presently being made, under the auspices of the National Science Foundation, of the feasibility of drilling a hole to the mantle in some oceanic area. If successful, this will provide positive identification of all layers.

Continental shelves. These are the submerged borders of the continents. The water depth is on the order of a few hundred meters, and the width of the shelf varies from a few miles in some places to a few hundred miles in others. The thickness of the crust is intermediate between that of continents and oceans, and its composition is continental. In some places, such as parts of the east coast of North America and South America, the continental shelf and continental slope are broad areas where erosion of the continental masses has resulted in the deposition of many thousands of meters of sediment during the past several million years. In other areas, notably the west coast of the Americas, there is only a narrow continental shelf, which does not receive much sediment

Submarine ranges and seamounts. These topographic features, of which the Mid-Atlantic Ridge and Bermuda are good examples, are unsurpassed in prominence anywhere on earth The ridges and isolated peaks are similar structurally. The tops and flanks are basaltic volcanic rock The cores, judging by the seismic velocity, appear to be a mixture of mantle and basaltic rock. Figure 10 shows a structure section across the Mid-Atlantic Ridge, based on seismic-refraction and gravity measurements The deep area near the center of the ridge is a great rift or crack. It has been noticed on many profiles running across the ridge and is apparently a characteristic feature of the entire Mid-Atlantic Ridge system The ridge is active seismically, and the belt of earthquake epicenters coincides with the axial rift, indicating a connection between the earthquakes and the formation of the rift.

Deep-sea trenches. Deep-sea trenches are important structural features associated with island arcs Notable examples are the Puerto Rico

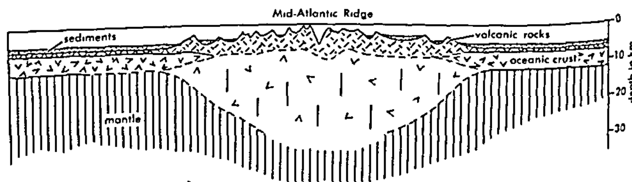


Fig 10. Structure section across the Mid-Atlantic Ridge. Vertical exaggeration X10

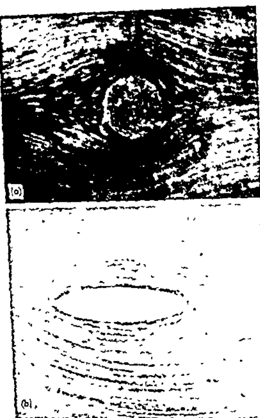


Fig 2 Flow at small Reynolds numbers (a) Around a circular cylinder (b) Around an elliptic cylinder (From L Prandtl and O. G Tietjens, *Applied Hydro- and Aeromechanics*, Dover, 1934)

motion between layers or between fluid and body surface In such a fluid the flow has little difficulty in moving from the leading edge to the point of maximum velocity, although there is some reduction in kinetic energy [illegible]

[illegible]ected to additional frictional contact with the surface, it eventually reaches a point on the aft surface at which its kinetic energy is completely dissipated, and it is drawn away from the surface by adjacent layers of fluid This action is known as separation and between the separation point and the trailing edge there is no flow parallel to the surface but rather a region containing irregular flow described as the wake For a sphere or cylinder, the transition from low pressure at the top radius to the rear stagnation pressure must take place in a relatively small path length, and the (adverse) pressure gradient along the surface is large (Fig 2a) If on the other hand the aft portion of the body is extended as in the usual subsonic airfoil, this pressure gradient is made much less severe, and although the length of surface over which frictional contact must take place is greater than in the case of the circular body, the separation point is much nearer the trailing edge, and the portion of aft surface subjected to the wake is less (Fig 2b).

The calculation of the resistance of a body to forward motion through a nonviscous fluid by superimposing the source, sink, and uniform stream flows results in a mathematical expression which graphically represents a body moving through a stationary fluid, or conversely, a stationary body immersed in a moving fluid. There is no flow separation, and the pressures at all points on the body exactly counterbalance, to give a net force of zero in all directions. *See* D'ALEMBERT'S PARADOX.

Drag. Because for the viscous case the flow never actually reaches the rear stagnation point, it cannot achieve the value of pressure predicted for it by the theory of a perfect fluid. Thus the pressure in the wake region is less than that at the front stagnation point, and a pressure differential exists in the streamwise or drag direction, resulting in a drag force. As the angularity of the airfoil increases positively relative to the flow (angle of attack increases), the front stagnation point moves farther down on the lower surface. Thus the flow passing from stagnation point to the trailing edge over the upper surface must have an increased velocity, while in the same way the lower surface velocity must decrease, and the lift force will therefore increase with angle of attack (Fig. 3a). As the angle of attack increases, the minimum pressure on the upper surface decreases, creating an increasing adverse pressure gradient over the aft portion of the upper surface, which in turn causes the point of flow separation to move farther for-



Fig 3 Flow pattern on an airfoil in a real fluid (a) Low angle of attack, unstalled, (b) high angle of attack, stalled (From P. P. Ewald, T Poschl, and L Prandtl, *Physics of Solids and Fluids*, Blackie, 1936)

Tectonic Interpretation, Geol. Soc. Am Spec. Paper 31, 1941; A. Wegener, *Origin of Continents and Oceans*, 1922.

Subsonic flight

Relative movement of a vehicle in air at a velocity appreciably below the velocity of sound. Subsonic flight extends from zero (hovering) to a speed of approximately 85% of the local speed of sound. At higher speeds, simplifications applicable to subsonic flight are no longer tolerable (*see* TRANSONIC FLIGHT). The type of vehicle may range from a small helicopter, which operates at all times in the lower range of the velocity scale, to an intercontinental ballistic missile which is operative throughout this and other velocity regimes, but is in subsonic flight for only a few seconds. The design of each is affected by the same principles of subsonic aerodynamics Subsonic flow of a fluid such as air may be further subdivided into a range of velocities in which the flow may be considered incompressible (below a velocity of approximately 300 mph) without appreciable error, and a higher range in which the compressible nature of the fluid becomes significant. In both cases the viscosity of the fluid is important. The theories which apply to incompressible, inviscid fluids may be used almost without modification in some low-subsonic problems, and in other cases the results offered by these theories may be modified to account for the effects of viscosity and compressibility. *See* COMPRESSIBLE FLOW; VISCOUS FLOW.

A typical subsonic wing cross section (airfoil) has a rounded front portion (leading edge) and a sharp rear portion (trailing edge) *See* AIRFOIL PROFILE. Air approaching the leading edge comes to rest at some point on the leading edge, with flow above this point proceeding around the upper airfoil surface to the trailing edge, and flow below passing along the lower surface to the same point, where the flow again theoretically has zero velocity. The two points of zero local velocity are known as stagnation points. If the path from front to rear stagnation point is longer along the upper surface than that along the lower surface, the mean velocity of flow along the upper surface must be greater than that along the lower surface. Thus in accordance with the principle of conservation of energy, the mean static pressure must be less on the upper surface than on the lower surface (*see* BERNOULLI'S THEOREM). This pressure difference applied to the surface area with proper regard to force direction gives a net lifting force (Lift is defined as a force perpendicular to the direction of fluid flow relative to the body, or more clearly perpendicular to the free-stream velocity vector)

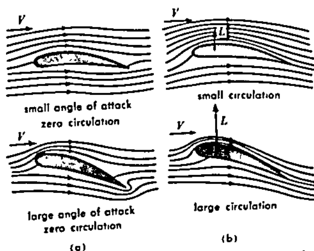


Fig 1. Flows around an airfoil (a) Without circulation. (b) With circulation

Necessity for circulation. For a body moving in a frictionless fluid with zero circulation, the flow attempts to align itself so that the rear stagnation point is equidistant from the front stagnation point along both upper and lower surfaces. For such a condition the rear stagnation point must move on the body as the angle of flow relative to the body changes. Thus if the body has a sharp trailing edge the local flow at the trailing edge must have infinite velocity as it moves from one surface to the stagnation point on the other surface (Fig 1a) Such a condition cannot exist in reality, and so there must be sufficient circulation (additional local velocity rearward along the upper surface) to move the stagnation point to the trailing edge (Fig 1*b*). Thus the flow from upper and lower surfaces must of necessity smoothly meet and leave the body at the trailing edge. This requirement is known as the Kutta condition, and when combined with the theory relating lift and circulation, independently developed by W. F Lanchester, an English engineer, W. M. Kutta, a German mathematician, and N. Joukowski, a Russian mathematician and scientist, will predict a lift force The Kutta Joukowski equation is given as $l = \rho U \Gamma$ where l = lift force per unit of wing span, ρ = mass density of the fluid (mass units per unit volume), U = linear velocity of fluid relative to the body, and Γ = fluid circulation (velocity times distance) The theory was originally applied to flow normal to the axis of a cylinder to explain the force produced by the rotation of the cylinder, or the Magnus effect. The fact that it predicts a reality even for spheres may be evidenced on any golf course or tennis court, and it is the basis for the entire circulation theory of lift for airfoil and wing

Effect of viscosity. In passing over the surface of a body, a frictionless fluid may be thought of as moving in layers which may slide relative to one another and relative to the surface with no retarding force. In moving from the front stagnation point to the upper surface, and hence to the trailing edge, the layers adjacent to the surface must travel from a region of high pressure (zero or low velocity) to a region of lower pressure (increased velocity), and then to a high pressure region at the trailing edge

Separation. In a flow containing no viscosity, the pressure at the rear stagnation point is exactly the same as that at the front stagnation point. In a viscous fluid the layers of flu resist any relat

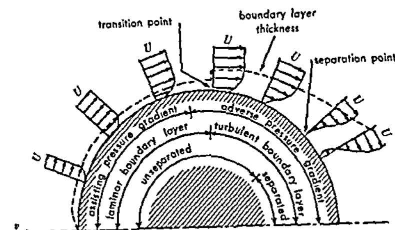


Fig 5 Diagrammatic representation of boundary-layer growth on a sphere for $R > 550{,}000$ The boundary-layer thickness is exaggerated for clarity. (From J. H. Dwinnell, *Principles of Aerodynamics*, McGraw Hill, 1949)

hand, separation of flow from the surface takes place in the boundary layer when the gradient of velocity is zero at the surface. The reduction in kinetic energy of the particles in the boundary layer due to adverse pressure gradient and constant surface impact has less effect on the turbulent boundary layer, with its rather full velocity profile, than on the laminar layer It is to be expected that the turbulent boundary layer will tend to delay separation, and this fact may be proven experimentally Because delayed separation is desirable for both drag reduction and increase in maximum lift (or delayed stall), it seems reasonable to desire high Reynolds number at least for the high angles of attack For low angles of attack, laminar flow is much more desirable

The designer has little control over the value of the Reynolds number, so he must resort to other means to achieve these aims. First, he may obtain low drag at the lower angles of attack by carefully shaping the airfoil so as to minimize adverse pressure gradient over the rear portion For such designs the drag usually rises rapidly at the high angles, however, the drag at high angles of attack is not considered disadvantageous, because the corresponding flight condition is take-off or landing, where the drag is not usually critical Nevertheless, the maximum lift obtainable from such shapes leaves much to be desired, and devices for increasing the lift must be used

Lift-control devices. Early researchers found that increased curvature of the thin externally braced wings caused an increase in lift with no increase in velocity or geometric angle of attack. The Wright airplane was controlled laterally by warping the proper wing. The weight and speed range of current designs requires a much thicker internally braced wing, and such warping is impractical. However, the camber (curvature of the line midway between upper and lower surface) of the airfoil may be changed by allowing a hinged section in the aft portion of the airfoil to rotate relative to the main structure, to produce a flap, aileron, elevator, or rudder, depending on the location of the particular airfoil in the aircraft. Each is a lift-producing device. These *trailing edge devices* tend to increase the adverse pressure gradient on the upper surface for a flap deflected downward, and on the lower surface for a flap deflected upward. One of the greatest problems in control-surface (hinged portion) design is that of maintaining flow over the upper surface of the deflected [illegible]

As the speed increases, the local velocity in the minimum-pressure region may become supersonic,

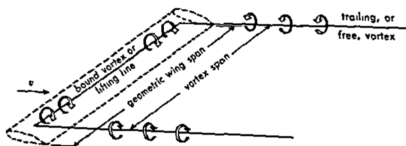


Fig 6 Representation of a rectangular wing by a horseshoe vortex

ward on the upper surface This forward movement has a dual effect on the characteristics of the airfoil. First, the wake region increases, thus exposing more of the surface to what has become an increased pressure differential in the stream direction, and therefore produces an increase in pressure or form drag. The second effect is found in the lift force, which is reduced by the premature separation of flow from the upper surface. Thus the lift force increases linearly with angle of attack until separation begins to occur on the upper surface forward of the trailing edge; then, as the angle of attack is increased further, the linearity disappears (slope of lift vs. angle of attack decreases) until at some angle (stall angle) the maximum lift is obtained (Fig. 3b). Further increase in angle of attack results in a loss of lift due to continued forward movement of the separation point. The increase in drag force with angle of attack is usually somewhat parabolic, depending on the particular airfoil, for all low subsonic velocities.

Aerodynamic force. If aerodynamic force F is assumed to be dependent upon such variables as fluid mass density ρ, relative velocity V of the body with respect to the fluid, a characteristic length l of the body, the compressible nature of the fluid (as symbolized by the speed of sound a), the viscous properties, as designated by the coefficient of viscosity μ, dimensional analysis reveals the following, where K is a nondimensional constant of proportionality, and α and β are experimental exponents

$$F = K\rho V^2 l^2 \left(\frac{a}{V}\right)^{\alpha} \left(\frac{\mu}{\rho V l}\right)^{\beta}$$

$$= C_F \frac{\rho}{2} V^2 S (M)^{-\alpha} (RN)^{-\beta}$$

The aerodynamicist prefers to work with forces in nondimensional form, and hence is interested in the force coefficient

$$C_F = \frac{F}{(\rho/2)V^2 S} (M)^{\alpha} (RN)^{\beta}$$

where F = force (as lift or drag), $(\rho/2)V^2$ = q = free-stream dynamic pressure, S = representative body area, M = free-stream Mach number, and RN = Reynolds number In the equation above it is implied that the force, and therefore the force coefficient, is dependent in an unspecified manner on both Mach number and Reynolds number (*see* MACH NUMBER; REYNOLDS NUMBER). The flow about the body, that is, the local velocity at all points in the flow, depends on the forces on the particles making up the flow. For low-speed conditions the only forces of any consequence are due to inertia and viscosity, and the ratio of inertia to viscous force is the Reynolds number. At higher speeds, where compressibility begins to be significant (above about 300 mph), there is introduced an elastic force, and the ratio of inertia force to elastic force is the Mach number. In the high subsonic region, both viscosity and compressibility play an important role, and the magnitudes of li and drag are dependent on each of the ratios. In order that two geometrically similar bodies m experience proportional forces (and therefore the same force coefficients) when placed at the sa attitude relative to a fluid (or to different fluids) the Reynolds number and Mach number for ea case must be identical (*see* DYNAMIC SIMILARITY). The requirement concerning Mach number does n hold for dynamic similarity at low speeds, but b comes more important than Reynolds number supersonic speeds. This principle is the basis f all experimental testing, whether it be in wind tu nels or in actual atmospheric conditions (s MODEL THEORY). To obtain the proper values the two ratios, it sometimes becomes necessary use test facilities which are pressurized or evac ated, heated or cooled, and gases which someti complicate the mechanical system.

The importance of the Reynolds number is gre est in the region of flow immediately adjacent the body surface (*see* BOUNDARY-LAYER FLO This layer of flow is thin compared to the size the body, and yet within it the flow velocity m increase from a required zero value on the surf to the local fluid velocity (Fig. 4). Evidence of s a layer is found in the film of dust which coll and remains on the surface of cars, trains, aircraft. Because of the usual thinness of the la the aerodynamicist usually feels justified in ass ing the flow outside the boundary layer to be viscid and subject to the laws of fluid mecha involving a perfect fluid. In this respect, the bou ary layer effectively increases the thickness slightly modifies the shape of a body (Fig. 5).

Effect of boundary layer. The drag on a b is a function of fluid friction and streamwise p sure differential, which in turn is a function of separation The fluid shearing stress on the b surface is a direct function of the velocity grad in the boundary layer, and the gradient is gr est in the turbulent boundary layer. On the o

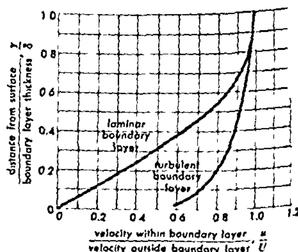


Fig 4 Dimensionless laminar and turbulent boundary-layer profiles.

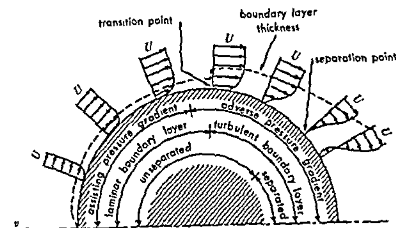


Fig 5 Diagrammatic representation of boundary-layer growth on a sphere for $R > 550{,}000$ The boundary-layer thickness is exaggerated for clarity (From J H. Dwinnell, *Principles of Aerodynamics*, McGraw-Hill, 1949)

hand, separation of flow from the surface takes place in the boundary layer when the gradient of velocity is zero at the surface The reduction in kinetic energy of the particles in the boundary layer due to adverse pressure gradient and constant surface impact has less effect on the turbulent boundary layer, with its rather full velocity profile, than on the laminar layer It is to be expected that the turbulent boundary layer will tend to delay *separation* and this fact may be proven experimentally Because delayed separation is desirable for both drag reduction and increase in maximum lift (or delayed stall) it seems reasonable to desire high Reynolds number, at least for the high angles of attack. For low angles of attack, laminar flow is much more desirable

The designer has little control over the value of the Reynolds number, so he must resort to other means to achieve these aims. First, he may obtain low drag at the lower angles of attack by carefully shaping the airfoil so as to minimize adverse pressure gradient over the rear portion For such designs, the drag usually rises rapidly at the high angles, however, the drag at high angles of attack is not considered disadvantageous, because the corresponding flight condition is take-off or landing, where the drag is not usually critical Nevertheless, the maximum lift obtainable from such shapes leaves much to be desired, and devices for increasing the lift must be used

Lift-control devices. Early researchers found that increased curvature of the thin externally braced wings caused an increase in lift with no increase in velocity or geometric angle of attack. The Wright airplane was controlled laterally by warping the proper wing The weight and speed range of current designs requires a much thicker internally braced wing, and such warping is impractical. However, the camber (curvature of the line midway between upper and lower surface) of the airfoil may be changed by allowing a hinged section in the aft portion of the airfoil *to rotate relative* to the main structure, to produce a *flap, aileron, elevator, or rudder*, depending on the location of the particular airfoil in the aircraft Each is a lift-producing device These trailing edge devices tend to increase the adverse pressure gradient on the upper surface for a flap deflected downward, and on the lower surface for a flap deflected upward One of the ... [illegible] ... region may become supersonic,

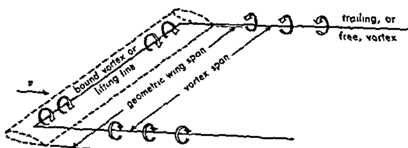


Fig 6 Representation of a rectangular wing by a horseshoe vortex

with the possible accompaniment of a normal shock wave farther aft on the surface, which usually interacts with the boundary layer and causes flow separation. *See* TRANSONIC FLIGHT.

Finite-wing effects. Because of the difference in pressure on upper and lower surfaces of a wing and because the wing is composed of airfoils, the fluid in the high-pressure region attempts to move around the tips and join the low-pressure fluid on the upper surface. The fluid moves rearward as it attempts to move upward, and the resultant path at the tip resembles a corkscrew. There is also a general outward flow on the lower surface and an inward flow on the upper surface which imparts a large amount of swirling motion to the fluid along the trailing edge as the upper and lower surface flows meet. Thus the flow downstream of the wing contains a sheet of vortices, ending in the large tip vortices. The large vortices are clearly visible in certain atmospheric conditions.

Lifting-line theory. The Prandtl lifting-line theory is based on the superposition of a field of square horseshoe-shaped vortex filaments which lie along the quarter-chord line of the wing and extend rearward to infinity, thus approximating the physical happenings described above (Fig. 6) The superimposed vortices along the lifting line of the wing provide a varying circulation which, when acted upon by a free-stream flow, produces a varying lift in accordance with the Kutta-Joukowski equation. Both the bound vortices and the trailing vortices behind the lifting line produce induced velocities known as downwash or upwash in the region surrounding the wing Thus there will be found by this theory an upward flow forward of the wing, a downward flow at the lifting line, and a downward flow behind the wing. These velocities are added vectorially to the free-stream velocity to obtain local flow magnitude and inclination in the regions mentioned. Along the lifting line the local angle of attack is reduced by varying amounts along the wing depending on the planform, but in general most highly tapered wings experience a greater reduction near the center line than at the tips. As the aspect ratio (ratio of wing span to mean chord) decreases, the downwash increases, thus requiring a larger angle of attack to produce the same wing lift coefficient. The angular reduction is known as the induced angle of attack The resultant velocity vector, representing the vector sum of the free-stream velocity and the downwash velocity, is larger than the free-stream velocity and is inclined downward relative to it; thus the resultant wing lift will be inclined rearward relative to the location of the lift vector for the wing of infinite aspect ratio, which has no downwash The component of the resultant lift which is normal to the infinite aspect-ratio lift vector, and therefore in the drag direction, is known as the induced drag, or the drag due to lift. Because the downwash increases with decrease in aspect ratio, the induced drag increases with reduction in aspect ratio. It is also obviously a direct function of lift, so that to minimize drag at high lift coefficients, the wing aspect ratio should be large This fact presents a problem to the designer, because a slender wing is more difficult to design structurally than a short, stubby wing of the same area. Another problem which confronts the designer is the effect of local angle of attack variation along the wing near stall. For some wings, if the root section is placed at an angle near stall, the tip sections will probably be stalled, and if the tip sections are placed near stall the root sections will not develop their full lift capacity. Thus the designer usually is forced to twist the wing-tip sections downward (washout) relative to the root sections, either aerodynamically by decreasing camber with distance from the wing center line, or by physically twisting the wing during construction. Nearly all modern airplane wings incorporate some manner of twist to produce proper stall characteristics.

Stability and control. The wing, as the lifting device for the aircraft, whether it be fixed, as in the airplane, or rotating, as in the helicopter, is probably the most important aerodynamic part of an aircraft. However, stability and control characteristics of the subsonic airplane depend on the complete structure. Control is the ability of the airplane to rotate about any of the three mutually perpendicular axes meeting at its center of gravity. Static stability is the tendency of the airplane to return to its original flight attitude when disturbed by a moment about any of the axes. The axes chosen are usually along the fuselage, along the wing and normal to these two in the vertical (relative to the airplane) direction, and are designated x, y, and z respectively

Rotation about the x axis is termed roll, and is achieved by producing unbalanced lift on opposite sides of the wing by means of the ailerons, located near the wing tips, which deflect differentially, down for the upgoing wing. Stability about the x axis is produced by sloping the wings upward in the spanwise direction, or in the xz plane. The result is termed dihedral, with downward slope known as cathedral or anhedral.

Control about the y axis is known as pitch control, and is achieved by the production of a moment about the center of gravity due to a force in the z direction produced either by a horizontal tail aft of the center of gravity or a control surface forward of the center of gravity, known as a canard. Stability is achieved primarily by the horizontal tail, because wing lift is usually forward of the center of gravity, thus providing an unstable moment, and most fuselages are inherently unstable (in the tailless airplane, stability is achieved by sweeping the wing back and twisting it so as to place the center of lift behind the center of gravity.)

Motion about the z or yaw axis is produced by the moment about the center of gravity due to a force in the y direction on the vertical tail or [illegible]

This force is in turn produced by rudder deflection. Stability is afforded primarily by this same vertical surface, but both the fuselage and wing also tend to resist rotation, and thus contribute to directional stability. Motions about one axis usually produce sufficient aerodynamic and dynamic disturbance to induce motion about other axes. See FLIGHT CHARACTERISTICS [J.E MA.]

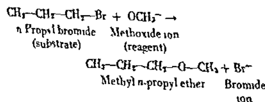
Bibliography: C. B. Millikan, *Aerodynamics of the Airplane*, 1941; C. D. Perkins and R. E. Hage, *Airplane Performance, Stability, and Control*, 1957; T. Von Kármán, *Aerodynamics*, 1954; K. D. Wood, *Technical Aerodynamics*, 2d ed., 1947.

Substitution reaction

One of a class of chemical reactions in which one atom or group (of atoms) replaces another atom or group in the structure of a molecule or ion. Usually, the new group takes the same structural position that was occupied by the group replaced.

Substitution reactions involve the attack of a reagent, which is the source of the new atom or group, on the substrate, the molecule or ion in which the replacement occurs. They involve the formation of a new bond and the breaking of an old bond. Substitution reactions are classified according to the nature of the reagent (electrophilic, nucleophilic, or radical) and according to the nature of the site of substitution (saturated carbon atom or aromatic carbon atom). See ELECTROPHILIC AND NUCLEOPHILIC REAGENT.

Systematic names for substitution reactions are composed of the parts: name of group introduced + de + name of group replaced + ation, with suitable elision or change of vowels for euphony. Thus, the replacement of bromine by a methoxy group (see equation below) is called methoxy-deb...



$CH_3{-}CH_2{-}CH_2{-}Br + OCH_3^- \rightarrow$ n-Propyl bromide (substrate), Methoxide ion (reagent)

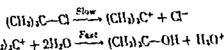
$CH_3{-}CH_2{-}CH_2{-}O{-}CH_3 + Br^-$ Methyl n-propyl ether, Bromide ion

Other reactive substrates include alkyl esters of sulfonic acids, *quaternary* ammonium salts, and tertiary sulfonium salts. Other effective reagents include mercaptide ions, halide ions, carbanions, water, and amines.

Th...

Th...

ste... ... the substrate into a carbonium ion and another fragment, and combination (usually rapid) of the carbonium ion with a nucleophilic reagent. An example is the hydrolysis of *tert*-butyl chloride:

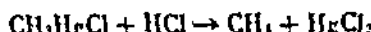


$$(CH_3)_3C{-}Cl \xrightarrow{Slow} (CH_3)_3C^+ + Cl^-$$

$$(CH_3)_3C^+ + 2H_2O \xrightarrow{Fast} (CH_3)_3C{-}OH + H_3O^+$$

This mechanism is favored when the substrate can yield a carbonium ion of comparatively low energy. The bimolecular or S_N2 mechanism involves one step in which formation of the new bond is simultaneous with breaking of the old bond. This mechanism is favored when the reagent has strong nucleophilic character and when the site of substitution is easily accessible to attacking reagents. The above reaction of *n*-propyl bromide goes by the S_N2 mechanism. The mechanisms of some reactions are intermediate between S_N1 and S_N2.

Electrophilic substitution at saturated carbon. This comparatively minor class is exemplified by the acid cleavage of alkyl mercury compounds:

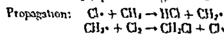


$$CH_3HgCl + HCl \rightarrow CH_4 + HgCl_2$$

Radical substitution at saturated carbon. Radical substitution usually involves a chain reaction. Such a reaction has three phases: initiation (in which radicals are generated), propagation (in which most of the actual reaction occurs), and termination (in which radicals are destroyed). These are illustrated by the mechanism of chlorination of methane:

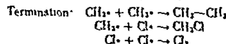


Initiation: Cl_2 + photon of light $\rightarrow 2Cl\cdot$



Propagation: $Cl\cdot + CH_4 \rightarrow HCl + CH_3\cdot$

$CH_3\cdot + Cl_2 \rightarrow CH_3Cl + Cl\cdot$



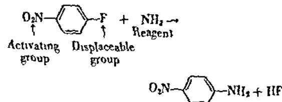
Termination: $CH_3\cdot + CH_3\cdot \rightarrow CH_3{-}CH_3$

$CH_3\cdot + Cl\cdot \rightarrow CH_3Cl$

$Cl\cdot + Cl\cdot \rightarrow Cl_2$

The propagation steps occur in a cyclic repetitive fashion: a product of one step is a vital *reactant* in another step. Hundreds of acts of propagation often occur for each act of initiation or *termination*.

Nucleophilic substitution at aromatic carbon. *An example* is the reaction of p-fluoronitrobenzene with ammonia:



... groups (—NR_3^+), the nitro group (—NO_2), and several others which are able to depart with an electron pair ... gen is selc... ally neces... *Effective* ... groups are those of strong electron-attracting character, such as —N_2^+, —NO_2, —SO_2CH_3, —$COCH_3$, and —CN. The hetero nitro-

gen atom in pyridine and related heterocycles is also a strong activating structure.

Most nucleophilic substitutions at aromatic carbon occur by a two-step intermediate complex mechanism. In the first step, the reagent becomes covalently attached to the carbon atom at the site of substitution to form a metastable intermediate complex. In the second step, the displaceable group is detached from the same carbon atom. Either formation of the intermediate complex or expulsion of the displaceable group may be the rate-limiting step. Some aromatic nucleophilic substitutions occur via an elimination-addition mechanism involving benzyne intermediates, and a few occur by the S_N1 mechanism. See BENZYNE.

Electrophilic substitution at aromatic carbon. The mechanism of this type of reaction usually involves two steps and an intermediate reaction complex: the reagent attaches to the carbon atom at the site of substitution, and the displaceable group then detaches. Hydrogen is the group most commonly displaced, but sulfo ($-SO_3H$), phosphono ($-PO_3H_2$), carboxyl ($-COOH$), mercuri ($-HgCl$, for example), and other groups may also be displaced. Many electrophilic reagents are effective; reactions are classified according to the group introduced, this being derived from the reagent. See DIAZOTIZATION; FRIEDEL-CRAFTS REACTION; HALOGENATION; NITRATION; SULFONATION.

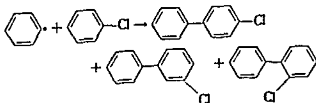
Electrophilic aromatic substitutions are activated (accelerated) by electron-releasing substituents such as



$$-OH, -O-CH_3, -NH_2, -NH-CO-CH_3,$$

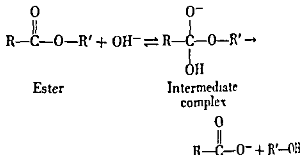
and $-CH_3$. Deactivating groups include $-NO_2$, $-SO_3H$, $-COOH$, and $-COCH_3$. Each type, activating or deactivating, acts most strongly on the ortho and para positions A common practical problem is to predict which of five displaceable hydrogen atoms in a monosubstituted benzene will be displaced by an electrophilic reagent. Activating groups such as $-OH$ are ortho-, para-directing (that is, they direct displacement of an ortho or para hydrogen). Deactivating groups, since they deactivate meta positions least, are meta-directing Halogen substituents, being mildly deactivating and yet ortho-, para-directing, constitute a special category.

Radical substitution at aromatic carbon. This also occurs by an intermediate complex mechanism, but details are not well understood. An example is the phenylation of *chlorobenzene:*



The phenyl radical may be obtained in various ways, such as by the decomposition of benzoyl peroxide or of N-nitrosoacetanilide by heat.

Nucleophilic substitution at carbonyl carbon. *Reactions representative of this type* are the hydrolysis of acid chlorides, acid anhydrides, and esters, or the reactions of these substrates with ammonia to form amides. These reactions usually occur by an intermediate complex mechanism, although in some cases the S_N1 mechanism prevails. Thus, common base-catalyzed ester hydrolysis (saponification) occurs as follows:



See ORGANIC CHEMICAL SYNTHESIS; ORGANIC REACTION MECHANISM. [J.F.B.]

Bibliography: J. F. Bunnett, Mechanism and reactivity in aromatic nucleophilic substitution reactions, *Quart. Revs. (London)*, 12:1–16, 1958; C. K. Ingold, *Structure and Mechanism in Organic Chemistry*, 1953; C. Walling, *Free Radicals in Solution*, 1957.

Subtilin

A mixture of polypeptidic antibiotics produced by *Bacillus subtilis*, American Type Culture Collection 6633. The antibiotic inhibits mainly the gram-positive bacteria. It is most active on dividing bacteria and on bacterial spores undergoing germination. Its antimicrobial activity on spores and on food-poisoning staphylococci has led to proposed uses in food processing

Subtilin is produced, in submerged culture, in a simple medium which is vigorously agitated. Sugar or a similar carbon source, an ammonium salt or certain organic nitrogen compounds, phosphate, sulfate, and ions of potassium, magnesium, iron, manganese, and zinc are required The metal requirements are generally more exacting for subtilin formation than for growth The fermentation time is 12 hours.

In young cultures, subtilin is associated with the *B subtilis* cells It is extractable by aqueous *n*-butanol; precipitable by dehydration of the butanol, by azeotropic distillation, or by addition of salt or *petroleum ether*, *and precipitable* from water by salt (see DISTILLATION) Subtilin A, the major component of the subtilin family, is separated by partition chromatography on silica gel or by liquid-liquid countercurrent distribution (see CHROMATOGRAPHY).

Subtilin A is a neutral compound with a molecular weight of about 3240; its structure is not known. On hydrolysis it gives the following amino acids (and number of each per molecule): alanine (1), L-aspartic acid (1), L-glutamic acid (3), glycine (2), L-isoleucine (1), *meso* lanthionine

(1), L-leucine (4), L-lysine (3), *beta*-methyllanthionine (4), L-phenyalinine (1), L-proline (1), sarcosine (2), tryptophan (1), L-valine (1). Those amino acids ordinarily not found in proteins have been found in hydrolysates of other antibiotics, sarcosine in the actinomycins and in etamycin, the lanthionines in nisin and in cinnamycin (*see* AMINO ACIDS).

Subtilin is resistant to inactivation by heat and weak acid, but it is inactivated by alkali. Methyl and some other esters of subtilin have enhanced activity, substitution on the free amino groups destroys activity Subtilin is largely inactivated by trypsin without extensive hydrolysis.

The mode of action of subtilin involves the bacterial cell membrane, causing leakage of essential cell constituents In this it resembles the antibiotic tyrocidine and the germicidal synthetic cationic *surface-active detergents*, such as cetyl trimethyl ammonium bromide (see ANTIMICROBIAL AGENTS; TYROTHRICIN). Spores subjected to severe heat while dormant, as in vegetable canning, become particularly susceptible on germination to subtilin and to the related antibiotic, nisin. This suggests a substantial lowering of the heat process requirements for certain foods. Practical tests are underway with tomato juice and with certain other vegetables

Subtilin is *nontoxic* when fed in rat diets It pre-

...OTIC, FOOD ENGINEERING, FOOD POISONING, BACTERIAL; INDUSTRIAL MICROBIOLOGY [I C L]

Subtraction

One of the four fundamental operations of arithmetic and algebra The first printed use of the symbol — to denote subtraction is in Johann Widman's *Behennde und hüpsche Rechnung*, Leipzig, 1489 Subtraction is often regarded as an operation inverse to addition that is, if a and b are numbers, the number $a - b$ is defined as that number which added to b gives a The more modern viewpoint eliminates subtraction completely by considering the number $a - b$ as the sum of a and that number (denoted by $-b$) which added to b gives 0 The number symbolized by $-b$ is called the inverse of b (with respect to addition) Every real number has a unique inverse (the number 0 is its own inverse) and so for each two numbers x, y the operation $x + (-y)$ gives a number. Clearly

$$[x + (-y)] + y = x + [(-y) + y] = x + 0 = x$$

and so the number $x + (-y)$ has the property of

...original ...tical operation greatly extended *See* ADDITION; ALGEBRA, DIVISION, MULTIPLICATION, NUMBER THEORY. [L.M BL.]

Succession, ecological

The study of changes in the community of organisms which occupy a given site, owing to increasing availability of environmental resources. Ecological succession is a gradual process brought about by the change in the number of individuals of each species and by the establishment of new species populations which may gradually crowd out the original inhabitants. Hence succession depends upon two main factors: (1) what organisms are in a position to invade a site, that is the floristic and faunistic resources of the general region; and (2) the rate of change of the habitat and its receptivity to these potential invaders. *See* POPULATION DISPERSAL.

Types of succession. Both allogenic and autogenic succession are recognized As a community exploits some resources of a habitat, leaving others untapped, physical and chemical changes are brought about. Organisms burrow, penetrating and aerating the soil, *changing the physical arrangement* of materials; they provide food for other organisms; they die, decay, and add to the variety of chemicals available. Inevitably, these activities modify the habitat. The resources needed by *one set of populations* become depleted, and conditions become favorable for the *invasion of others*; communities pave the way for one another in autogenic succession.

In instances where some external f... tal, ... river ... allogenic succession. In most successions both forces are at work. *See* BIOSPHERE, GEOCHEMISTRY OF; ECOSYSTEM, FOOD CHAIN.

Sere. The unoccupied habitat determines the nature of the ...

...ngus populations, decomposing tree stumps with

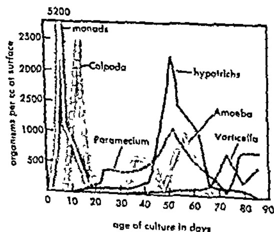


Fig. 1. Succession of protozoan populations in a hay infusion (After Woodruff, 1912, from E. P Odum, Fundamentals of Ecology, 2d ed, Saunders, 1959)

successive insect communities, hay infusions (Fig. 1), and the more extensive habitats, such as ponds, sections of rivers, oyster banks, mud flats, sand dunes, abandoned fields, lava flows, or other terrain types.

Those who are interested in population dynamics have found it convenient to start from a carefully controlled small habitat; others, interested in the natural revegetation of certain areas, have found it necessary to establish permanent quadrats, sample areas to be observed year after year. But succession in such quadrats tends to slow down after a few years as more and more perennial plants become established. It is then necessary to deduce the course of succession from a comparison with similar habitats, which were invaded at earlier times, and hence are farther advanced in succession. This procedure introduces a number of possible errors. The compared plots may have slightly different environmental conditions, and the availability of disseminules is never exactly the same in different places and at different times. Because of these difficulties, knowledge of succession is partly extrapolated from observed short-term changes and partly synthesized from isolated examples. In the case of some early American authors, personal desire for order led to the construction of such theories as the monoclimax concept which is considered later.

Classification of seres. It is convenient to classify successional sequences on the basis of the starting habitat. Hydroseres, or hyrarch successions, are those in which pioneer plants invade *open water, eventually forming some kind of soil* such as peat or muck, the Verlandung of German ecologists. Xeroseres, or xerarch successions, are initiated on dry ground such as rock, sand, clay or similar sterile locations. Such subdivisions as psammosere (shifting sand) and lithosere (rock) are also useful.

Another important distinction is based upon whether the open habitat has never before been occupied, thereby forming the start for the prisere (primary succession), or whether the habitat has been cleared artificially, leaving vestiges of soil, seeds, and debris from previous occupancy and starting a subsere (secondary succession)

Stages of succession. As an area is invaded and taken over by successive populations of plants and animals, the physiognomy of the vegetation changes. Initially isolated pioneer plants eventually give way to a closed carpet (Fig. 2) bringing about consolidation of the plant cover. As the habitat is further exploited, it is taken over by new species having a life form which can more fully utilize its resources (subclimax stage) until a community is finally established which perpetuates itself indefinitely because each species reproduces and maintains a stable number of individuals. Such stable, terminal communities are called climax communities.

This development of a stable community has been compared by F. Clements with the development of an organism—an analogy which is too

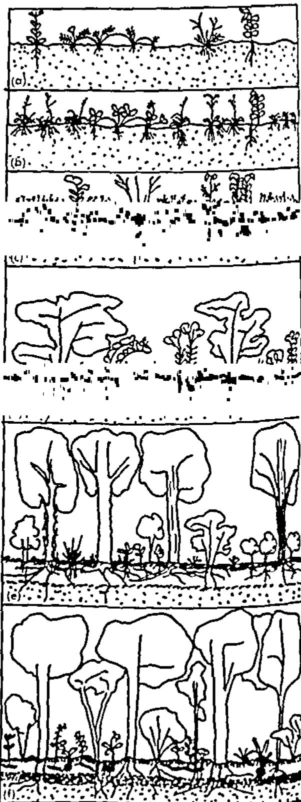


Fig. 2. Plant succession on abandoned field in deciduous forest region. (*a*) Early pioneer stage on azonal soil. (*b*) Late pioneer stage of poverty grass. (*c*) Consolidation stage of Kentucky bluegrass and goldenrod. (*d*) Late consolidation stage of hawthorn, blackberry, and asters on intrazonal soil. (*e*) Subclimax stage of hickory, oaks, and elm. (*f*) Climax stage of sugar maple, beech, and tulip poplar, with rich spring flora on zonal soil.

far fetched to have much significance because succession does not proceed in the same genetically determined fashion, and the end result is not anything like an organism in its degree of functional integration and structural stability Yet this superorganism concept occurs repeatedly in the literature

NATURE OF SUCCESSION

On the basis of physiognomy, it is possible to divide succession into pioneer, consolidation, subclimax, and climax stages (Fig 2). The objection has been raised that such stepwise transitions are not borne out by close study of the populations There is a continuously changing array of species, and any subdivision of such a continuum must be an arbitrary decision based on expediency or individual taste Yet the question of the objective reality of communities in successional sequences has not been settled to the satisfaction of all ecologists; the kind of data which would settle this dispute requires field observation over many decades.

It may be well, however, to keep in mind the following characteristics of ecological succession listed by R. Whittaker in 1953: (1) its nature is continuous; (2) comparable habitats are not always settled at the same rate or by the same group of species; (3) certain stages may be prolonged, telescoped into one another, or omitted entirely in different habitats or in different parts of the same habitat

Succession-retrogression cycles. The progress of succession is often interrupted by natural or man-made disturbances which open up closed communities, clear much of the habitat, or kill off certain key species of the community If this disturbance continues, the vegetation will adjust permanently and stabilize in a kind of disturbance-controlled climax (disclimax) The Irish and German heath lds are maintained by sheep grazing and burn g of the sod; the palmetto-pine lands of the utheastern United States are controlled by forest es

If the disturbance does not recur succession sim iv receives a setback (retrogression) and starts gain from an earlier stage Several examples of gular al
a cyc
lly succ
the ca ... bogs, dunes, heaths, and certain arctic areas where the permafrost level fluctuates.

Mosaic nature of succession Succession proceeds at different rates in different parts of the habitat New invaders, at first randomly occupying small spots, soon become centers for the establishment of clusters of new species, because of better shading, mi...
ers and in
ready estab ... a mosaic of expanding and contracting pieces of vegetation comes about It may persist into the climax where gaps in the tree canopy, slight undulations of the terrain, and other similar phenomena are responsible for the pattern.

Succession and zonation. Succession frequently progresses from the edge to the center of an open habitat, creating belts of vegetation. The belts in the center of such areas are still in the early stages of succession while the edge has progressed to a later stage. Thus, communities which will succeed one another in time become laid out in spatial arrangements. This makes possible the study of successional stages simultaneously, with transects of sample areas across the zones of vegetation (Fig 3).

Zonation, however, is not necessarily evidence of succession; nor do the communities found in successive zones necessarily reflect the true course of succession. On a river bank with a zonation of willows, cottonwoods, sycamores, ashes, and maples, each species occupies a slightly different site, with different degrees of flooding and soil deposition. Hence they are not successive stages in a sere, but merely adjacent stands of trees in different habitats.

Convergence toward the climax. A key observation about the process of succession is that the initial stages of various hydroseres and xeroseres may differ greatly in appearance and species composition, but as succession proceeds, later stages bear a progressively greater resemblance to one another, first in their physiognomy, and also later in species composition

For example, lichens and mosses are the common invaders of rock habitats, whereas various annuals and grasses settle dry sand plains; eventually, however, both habitats will support some kind of oak woodland. Although these trees may not be the definitive climax of these sites, they nevertheless demonstrate a convergence towards a certain type of vegetation. This convergence is largely dependent upon the development of the soil, and research on succession must be firmly based upon a study of soil in different stages.

Succession based upon soil development. In most primary successions the initial substrate is not differentiated into soil horizons (azonal soil). In



Fig 3 Zonation of vegetation in the succession on open sand flat with jack pine as subclimax

successive insect communities, hay infusions (Fig. 1), and the more extensive habitats, such as ponds, sections of rivers, oyster banks, mud flats, sand dunes, abandoned fields, lava flows, or other terrain types.

Those who are interested in population dynamics have found it convenient to start from a carefully controlled small habitat; others, interested in the natural revegetation of certain areas, have found it necessary to establish permanent quadrats, sample areas to be observed year after year. But succession in such quadrats tends to slow down after a few years as more and more perennial plants become established. It is then necessary to deduce the course of succession from a comparison with similar habitats, which were invaded at earlier times, and hence are farther advanced in succession. This procedure introduces a number of possible errors. The compared plots may have slightly different environmental conditions, and the availability of disseminules is never exactly the same in different places and at different times. Because of these difficulties, knowledge of succession is partly extrapolated from observed short-term changes and partly synthesized from isolated examples. In the case of some early American authors, personal desire for order led to the construction of such theories as the monoclimax concept which is considered later.

Classification of seres. It is convenient to classify successional sequences on the basis of the starting habitat. Hydroseres, or hyrarch successions, are those in which pioneer plants invade open water, eventually forming some kind of soil such as peat or muck, the Verlandung of German ecologists. Xeroseres, or xerarch successions, are initiated on dry ground such as rock, sand, clay or similar sterile locations. Such subdivisions as psammosere (shifting sand) and lithosere (rock) are also useful

Another important distinction is based upon whether the open habitat has never before been occupied, thereby forming the start for the prisere (primary succession), or whether the habitat has been cleared artificially, leaving vestiges of soil, seeds, and debris from previous occupancy and starting a subsere (secondary succession)

Stages of succession. As an area is invaded and taken over by successive populations of plants and animals, the physiognomy of the vegetation changes Initially isolated pioneer plants eventually give way to a closed carpet (Fig. 2) bringing about consolidation of the plant cover. As the habitat is further exploited, it is taken over by new species having a life form which can more fully utilize its resources (subclimax stage) until a community is finally established which perpetuates itself indefinitely because each species reproduces and maintains a stable number of individuals. Such stable, terminal communities are called climax communities.

This development of a stable community has been compared by F. Clements with the development of an organism—an analogy which is too

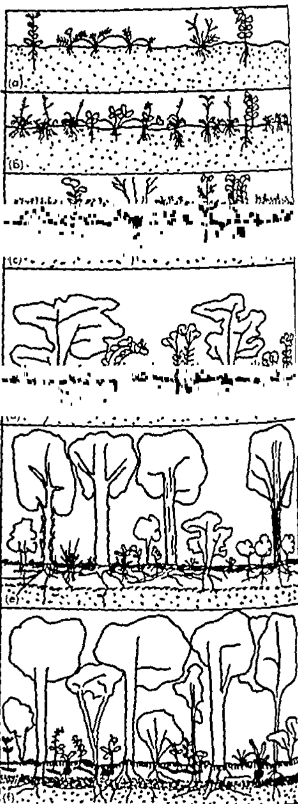


Fig. 2. Plant succession on abandoned field in deciduous forest region. (a) Early pioneer stage on azonal soil. (b) Late pioneer stage of poverty grass (c) Consolidation stage of Kentucky bluegrass and goldenrod. (d) Late consolidation stage of hawthorn, blackberry, and asters on intrazonal soil. (e) Subclimax stage of hickory, oaks, and elm. (f) Climax stage of sugar maple, beech, and tulip poplar, with rich spring flora on zonal soil.

the Northern Hemisphere has been in continuous change as a result of increasing mildness of climate. Such changes are not commonly regarded as successions, for they are not the result of expanding usage of the habitat, and proceed much more slowly than successions.

When climate changes, a change in the adjustment of the vegetation to its habitat results. This is especially true in the higher stages of succession; the pioneer stage is limited chiefly by the extreme nature of the substrate, not directly by climate. In the climax complex, however, a shift towards dry climate forces the trees to rely more heavily upon water stored in the soil (edaphic compensation); hence, they survive only in protected coves. Under these circumstances the upland is taken over by another community. A drought-resistant topographic climax now becomes climatic climax, while the former climax leads a precarious existence in sheltered places. Such relic climaxes, typical of climates which are cooler and more moist than the general [illegible]

the climax is also changed on the upland, and remains only on dry, exposed sites, where it persists as a preclimax. In the sugar maple-beech forests of the Middle West, enclaves of oak woodland are preclimax communities, whereas coves of hemlock forest are postclimax. *See* PALYNOLOGY.

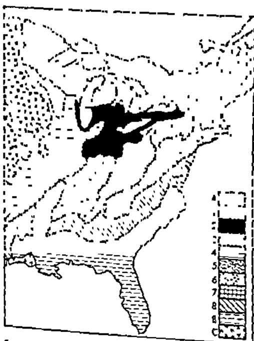


Fig. 4. Climax regions within the deciduous forest formation. 1, hemlock, northern hardwoods region, 2, beech-maple region, 3, maple-basswood region, 4, oak-hickory region; 5, western mesophytic region (tension area), 6, mixed mesophytic region, 7, oak-chestnut region, 8, oak-pine region, A, boreal needleleaf forest formation; B, broadleaf evergreen forest formation; C, the grassland formation. (Redrawn after E. L. Braun, 1950)

Tension zones. Attempts to map the extent of regional climaxes, such as those of E. L. Braun, are usually hampered by two problems inherent in transition zones (Fig. 4): continuity of the vegetation, and interpenetration of climaxes. Unless some feature of the topography, such as a mountain range or wet lowland, forms the abrupt border of the area of each of the climax species, there is bound to be so much overlapping and blending between two climax communities that it is impossible to draw the line. This has led certain workers to recognize only individual stands of vegetation not classifiable into abstract communities. Such views do not encourage vegetation mapping, unless one uses blending colors.

On the edges of each regional climax the role of certain post- or preclimax communities becomes increasingly important because here they are still close to their own climate. The result is an interpenetration of climax communities on the upland. The question as to which of these is entitled to the name regional climax becomes completely impractical where much of the land is contested by two or three climaxes.

The result is that in certain areas the polyclimax concept has great merits, especially in tension zones, in hilly terrain, and in areas with rapidly changing climates; in others, the monoclimax concept is a better model, especially in large uniform areas with a history of stable climate.

THE IMPORTANCE OF SUCCESSION

Prevalence of succession. Many atlases present maps of vegetational formations (biomes) of the world. A single formation, such as the deciduous forest, would contain a large number of regional climaxes (Fig. 4), each accompanied by sub-, post-, and preclimaxes. Inevitably some of the area, often much of it, is given to early stages of succession and cannot be mapped on a world-wide scale. *See* BIOME.

Ecological succession was discovered in the temperate zone because much of the land is occasionally cleared, and vegetation invades the habitat in conspicuously different stages. To the north, in the tundra and taiga, there is also much disturbance, especially by ice and seasonal upheavals of the soil. These disturbances are so continuous and the growing season is so short, however, that succession is much less important as a force in the vegetation than is the response to substrate, length of snow cover, and other factors.

In the moist tropics, large areas are relatively undisturbed. Here the vegetation is continuously in climax condition (sometimes edaphic climax). Where disturbances occur, the initial stages of succession proceed rapidly after the land is stripped. However, the forest does not immediately return to climax condition, and for a time palms, tree ferns, and vines form a junglelike subclimax.

the course of succession, the pioneer and consolidation plants take up certain minerals and deposit organic matter on the surface, thus starting the process of soil development. Depending upon rainfall, temperature, and soil organisms (microbes, fungi, insects, larvae, worms) the decayed materials infiltrate the soil to certain depths, where they are deposited and reutilized. If the initial substrate is rocky, plant roots and rhizoids may break it up while their stems and leaves trap dust and sand particles. If the initial substrate is water, an organic layer may form near the surface or on the bottom, which in due time becomes thick enough to support terrestrial plants and animals. Gradually a layered soil is formed (intrazonal soil).

Two aspects of the soil need further mention: (1) It has been stated that each successive stage of a sere makes more extensive use of the resources of the habitat. This implies that the climax is the one community which uses the most soil minerals in the shortest time, returning them to the soil as the leaves fall or as the organisms die. The increasing efficiency of successive communities, although unconfirmed, is now being investigated by those interested in the transfer of matter and energy in ecosystems. (2) It has also been mentioned that a convergence of the plant cover is evident as succession progresses. The question of whether there is also a soil climax and whether it coincides with the biological climax has a twofold answer. Each particular substrate has its own maximum degree of differentiation, and is considered mature (zonal soil) when this maximum is reached This condition is generally achieved under climax conditions of the vegetation which it supports. A stable soil accompanies a stable plant cover. However, there are certain unalterable properties peculiar to each soil. Within the time needed for a succession, the mineral composition of the soil changes only slightly. A clay soil will always be a clay soil; a soil deficient in calcium will not improve no matter how many plants grow on it. Only if calcium is present in layers below the topsoil can plants bring it up and improve the topsoil. Although the texture of soil may be slightly modified by additional organic matter, its basic properties are not changed. For this reason, the convergence toward the climax is never complete. Mature zonal soils, although similar in degree of development, remain as different as their parent materials in mineral make-up *See* Soil.

THE CLIMAX COMPLEX

Ecological succession eventually leads to a community with greater stability and permanence than the successional stages However, there may be many reasons why succession comes to an end. Repeated fires, drainage patterns, and topographic factors may all be responsible, and within one region several communities of similar physiognomy, but with different species composition, make up a complex of climaxes. Only a few commonly recognized types of climax can be discussed here. In no other aspect of ecology has nomenclature been so rampant.

Subclimax. The subclimax is the first step in converting the consolidation stage to climax community. In forest regions, this happens when tree seedlings become established in a meadow and create areas of shade These trees generally are not the final climax species, for often they will only germinate and grow in the open, not in dense shade. Thus the subclimax resembles the climax in structure, but it differs in species composition Both could be deciduous forest, or in another area both might be grassland.

Usually, the habitat of the subclimax limits further progress. Blocked drainage in swamp forests; coarse, porous soils in pine barrens; and seasonal flooding of river forests are examples of such limiting factors. If this is the case, the subclimax takes on the permanent role of edaphic or topographic climax.

Regional climax. This climax develops from subclimax communities in habitats where drainage is good but not excessive, so that no topographic or edaphic inhibitions exist. At this stage the interdependence of species is greatest, with growth rhythms, life forms, and local distribution patterns arranged so that the most efficient use is made of the available resources. Whether the regional climax is the most efficient and productive community of its successional sequence will remain a matter of speculation until better productivity measurements are possible.

According to some authors, this community has freed itself from its edaphic and topographic controls and is limited only by climate, hence the synonym climatic climax

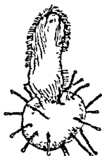
Monoclimax and polyclimax concept. F. Clements' original concept of succession was that all seres would end in the climatic climax; hence, the regional climax would be identical for all seres. This is the monoclimax concept. This concept has not stood the test of field observation. Although it is true that many climax species can grow on mature zonal soils regardless of whether they originated from limestone, sand, or peat, it is an exaggeration to say that soil has no influence at all on the composition of the climax.

The alternative polyclimax concept takes note of four factors: (1) the effect of local relief, creating a mosaic rather than a homogenized plant cover, (2) topographic differences responsible for different topographic climaxes, on various exposures on mountains; (3) soil inhibitions creating edaphic climaxes; and (4) climatic changes, leaving isolated relic communities as climaxes of the past. As a result of all this, very little of the landscape may actually support the regional climax community

CLIMAX AND CLIMATE

Effects of changing climate. It is known from the analysis of pollen grains found in lake deposits that for many thousands of years the vegetation of

ized features Most conspicuous are their tentacles, often many in number, which serve as mouths. These multiple organelles of ingestion become fastened to the pellicle of prey organisms, generally passing ciliates. By forces still not entirely understood, they are used to suck out the prey's protoplasm to provide sustenance for the suctorian's own body Nearly all species are stalked, and the sedentary, mature forms are devoid of any external ciliature Young larval forms are produced by both endogenous and exogenous budding These forms bear locomotor cilia and serve, as in the case of the peritrichs, for dissemination (see illustration).

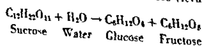


Endogenous budding in the suctorian *Podophrya*, a species which measures 10–28 μ

From an evolutionary and phylogenetic point of view, it is very significant that an infraciliature is present throughout the curious suctorian life cycle Silver staining techniques, in particular reveal it just as clearly in the adults without cilia as in the ciliated larval forms *Acineta* is a common species *Podophrya*, shortly after the emergence of the young through the birth-pore, is an example of endogenous budding *Ephelota* exhibits multiple exogenous budding as its means of asexual reproduction *See* HOLOTRICHA. [J O C.]

Sugar

The word "sugar," when unmodified, refers to sucrose the common sugar of commerce. It is a disaccharide of the formula $C_{12}H_{22}O_{11}$ which is split by hydrolysis into two monosaccharides, or simple sugars glucose (dextrose) and fructose (levulose).



$$\underset{\text{Sucrose}}{C_{12}H_{22}O_{11}} + \underset{\text{Water}}{H_2O} \rightarrow \underset{\text{Glucose}}{C_6H_{12}O_6} + \underset{\text{Fructose}}{C_6H_{12}O_6}$$

Sucrose rotates the plane of polarized light to the right, as does glucose but fructose is so strongly levorotatory that it overcomes the effect of glucose Thus mixtures of equal amounts of glucose and fructose are levorotatory The hydrolytic reaction is called inversion of sugar, and the product is invert sugar or, simply invert *See* CARBOHYDRATE, OPTICAL ACTIVITY

Sucrose is widely distributed in nature, having been found in all green plants which have been carefully examined for its presence. The total quantity of all sugars formed on earth each year has been estimated at a colossal 400,000,000,000 tons. Commercial sugar of fairly high purity, which has been put through centrifugals to separate molasses, totals about 50,000,000 tons a year, and the amount is rapidly increasing.

In spite of its availability in all green plants, sucrose is obtained commercially in substantial amounts from only two plants: sugar cane, which supplies about 60% of the world total, and the sugar beet, which provides 40%.

Cane sugar manufacture. The manufacture of cane sugar is usually done in two series of operations. First, raw sugar of about 98% purity is produced at a location adjacent to the cane fields. The raw sugar is then shipped to refineries, where a purity close to 100% is achieved.

Raw cane sugar. The production of raw cane sugar begins with growing the cane in tropical or subtropical areas (see SUGAR CANE). The cane is harvested after a season which varies from seven months in subtropical areas to 12–22 months in the tropics The cane stalks are harvested either by heavy handknives, or mechanically. The trend is toward mechanization The stalks are transported to a mill by oxcart, rail, or truck where they are crushed and macerated between heavy grooved iron rolls while being sprayed with water countercurrently to dilute the residual juice. The expressed juice contains 95% or more of the sucrose present. The fibrous residue, or bagasse, is usually burned under the boilers, although increasing amounts are being made into paper, insulating board, and hard board, as well as furfural, which is an intermediate for the synthesis of nylon 66

Cane juice. The cane juice is treated with lime [illegible] of crystallizable sugar. The juice is then heated to facilitate the precipitation of impurities, which are then removed by continuous filtration The purified juice is concentrated by multiple stage vacuum evaporation (usually four or five stages) and when sufficiently concentrated is boiled to grain or seeded with sucrose



Interior view of raw sugar mill. Heavy rollers squeeze sugar-bearing juices from macerated cane (Sugar Research Foundation, Inc)

It is not coincidental that succession was studied most intensely in the area where it is most neatly laid out in stages, the deciduous forests of Europe and America.

Uses of succession by man. The activities of man in the landscape generally have the effect of stopping succession at an early stage. Fire was used by most primitive people to keep forests suitable for hunting, or to clear land for primitive agriculture. Since almost all major crop plants are annuals of the pioneer stage, regular plowing is needed to hold down succession. Other methods of checking succession include mowing and grazing, which favor sod-forming grasses and kill most tree seedlings; and selective weed control with chemicals which generally damage broad-leaved plants more than grasses and the like.

By manipulation of the drainage pattern of the land, succession can be either retarded or accelerated depending upon the owner's intended purpose. Such manipulation may be used to provide good fishing conditions, or to establish a cranberry bog, a cow pasture, or a stand of productive timber. *See* CONSERVATION OF RESOURCES; LAND USE PLANNING. [K LE]

Bibliography: W. C Allee, A. E. Emerson et al., *Principles of Animal Ecology*, 1949; B W Allred and E. S. Clements (eds), *Dynamics of Vegetation*, 1949; E. L. Braun, *Deciduous Forests of Eastern North America*, 1950; F. E. Clements, *Plant Succession and Indicators*, 1928; H A. Gleason, The individualistic concept of the plant association, *Amer. Midland Naturalist*, 21:92–110, 1939; E. P Odum, *Fundamentals of Ecology*, 2d ed , 1959; J. E. Weaver and F. E. Clements, *Plant Ecology*, 1929.

Sucker

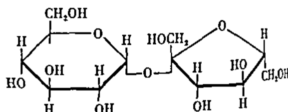
Any fish belonging to the family Catostomidae, totaling about 100 species, all North American except for one in Siberia and one in China. Suckers are soft-finned fishes, with cycloid scales and a protrusible mouth, usually ventrally located The suckers, especially the smaller members of each family, closely resemble the minnows. In addition to the many fishes called suckers, this family includes the buffalo fishes, quillbacks, and redhorses. The typical sucker, such as the redhorse or chubsucker, can be distinguished by the thickened lips, which are well developed on most of the family. Suckers are bottom feeders, feeding by suction and eating both plants and animals They are valuable as forage fish when small, and the larger species are of some commercial importance. The buffalo fishes alone contribute from 15,000,000 to 20,000,000 lb to the commercial harvest annually. *See* CYPRINIFORMES. [J.D.B]



The common sucker, Catostomus commersonnii; length to 28 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Sucrose

An oligosaccharide also known as saccharose, cane sugar, beet sugar, α-D-glucopyranosyl-β-D-fructofuranoside. The sugar occurs universally throughout the plant kindgom in fruits, seeds, flowers, and roots of plants. Honey consists principally of sucrose and its hydrolysis products. Sugar cane and sugar beets are the chief sources for the preparation of sucrose on a large scale. Another source of commercial interest is the sap of maple trees The consumption of sucrose in the United States is more than 100 lb per capita per year *See* OLIGOSACCHARIDE.

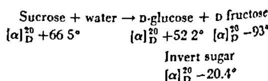


Sucrose

The specific rotation of sucrose $[\alpha]_D^{20}$ is +66.5° and melting point (mp) 186°C. It is nonreducing and is not fermentable by yeast. Sucrose is very soluble in water and crystallizes from that medium in the anhydrous form. *See* OPTICAL ACTIVITY.

The bacterium *Pseudomonas saccharophila* contains an enzyme, sucrose phosphorylase, which catalyzes the synthesis of sucrose from α-D-glucose-1-phosphate and D-fructose. In plants, sucrose is synthesized from uridine diphosphate D-glucose and D-fructose or D-fructose-6-phosphate. In the latter case, the resulting sucrose phosphate is hydrolyzed by an enzyme, phosphatase, to yield free sucrose *See* ENZYME.

The specific rotation of sucrose is +66 5° in water; it is readily hydrolyzed by acids and by the specific enzyme sucrase (invertase) to yield equal amounts of D-glucose and D-fructose.



$$\underset{[\alpha]_D^{20}\ +66.5^\circ}{\text{Sucrose}} + \text{water} \rightarrow \underset{[\alpha]_D^{20}\ +52.2^\circ}{\text{D-glucose}} + \underset{[\alpha]_D^{20}\ -93^\circ}{\text{D-fructose}}$$

Invert sugar $[\alpha]_D^{20}$ −20.4°

As the levorotation due to D-fructose exceeds the dextrorotation due to D-glucose, the hydrolysis results in a change in sign The process is therefore called inversion, and the equimolar mixture of the two sugars is known as invert sugar. [W.Z H]

Suctorida

An order of the Holotricha whose members were long considered quite separate from "true" ciliates. These forms show a num r of highly special-

The total quantity of sucrose obtained from waste fruit is statistically negligible. On the other hand, a substantial fraction of the total sucrose consumed is that naturally present in a very large number of fruits, vegetables, and nuts.

Lactose. Cow's milk on a dry basis is about 38% lactose or milk sugar. In the United States about 13 lb of lactose per capita per annum are consumed in milk products. This compares to 96 lb for sucrose. When milk is converted into cheese, the lac-

crose and one-sixth to one-half as sweet, depending on concentration. Uses are actively being sought.

Starch. The corn products industry is discussed elsewhere, but it should be mentioned that

tose, maltose

of amylases on starch. This hydrolysis has been of great importance, for thousands of years, in splitting starches for alcoholic fermentation. As yet, there is no large-scale production of pure maltose.

Maple sugar. When America was discovered by the white man, the Indians were collecting and concentrating the juice of the hard maple (*Acer saccharum*), thus making maple syrup. The practice was quickly accepted by the new settlers and has been an industry ever since in the regions where hard maples are common, principally the northeastern United States. Recent research has disclosed the curious fact that the maple flavor does not exist in the sap but is developed by heating it. By additional heating at about 120°C, a flavor 4–5 times more intense can be developed. Maple

licious flavor, delight of gourmets, makes up only a small per cent. Fairly satisfactory imitation maple flavors are available.

Honey. Honey is a form of relatively pure invert sugar dissolved in water to form a concentrated solution, but also containing precious flavors derived from the nectar of the flowers from which it was obtained by the bee. Nutritionally, it is nearly equivalent to invert sugar but contains an excess of fructose over glucose. The sucrose found in the flowers is inverted by the enzyme, honey invertase. Tupelo honey is remarkable in containing about twice as much fructose as glucose, and hence has little tendency to deposit glucose crystals. The ready availability of the food energy in honey was known to athletes in ancient Greece. Only in recent times has it been discovered that, paradoxically, the energy of sucrose is still more quickly available. The flavors of various honeys run a wide gamut. That from the Mt. Hymettus region, which is flavored by wild thyme, has been known and treasured since the poems of Homer.

Molasses. Virtually all molasses is distributed in the form of a concentrated viscous solution, but it can be reduced to a powder by means of spray drying. It can then be handled without an investment by the customer in tanks, pipes, and pumps. The problem is not so much the evaporation, as preventing later contact with moisture which converts it to a gummy mass. The availability of vapor-proof bags, for example, those lined with polyethylene, has provided one solution to the problem. There are also various additives which, when mixed with molasses, reduce its tendency to pick up moisture. So far (1960) dried molasses has made little headway against the standard practice of handling the concentrated solutions.

Comparison with synthetic sweeteners. Sug-

sucrose. After dulcin and certain others had failed to win approval by the U.S. Food and Drug Administration, the salts of cyclohexylsulfamic acid were permitted under various trade names. These are only about 30 times as sweet as sucrose, but lack the bitterness which many people observe in saccharin. Cyclamates are also more resistant than saccharin to the action of hot water and therefore find use in canning fruit with lower caloric content than that sweetened with sucrose. Tests have shown that such fruit is much preferred to that canned in plain water but not equal in flavor to a sucrose pack. About 1% of the sweetener market goes into synthetics.

Syrups. Syrups are relatively concentrated, somewhat viscous, solutions of various sugars, frequently in admixture to hinder crystallization. The Dutch word for syrup is *stroop*, which, somewhat altered, has entered our language in the term "blackstrap molasses."

Approximately 11% of the sucrose sold in the United States (1958) is in the form of syrups of high purity. These so-called "liquid sugars" have the double advantage of economy of handling, being moved with pumps, pipes, tank trucks, and tank cars, and also have a high degree of sanitation, since closed containers are used. It is necessary, however, to pay freight on the water content, which limits the distance between a refinery and a liquid sugar user. This problem has been met in two ways: First, some sugar manufacturers have established centers for the distribution of liquid sugars at places remote from the refinery. Granulated sugar is transported to the distribution center in a densely populated area and there dissolved in water and delivered to customers. Second, granulated sugar may be distributed to a customer in tank cars and dissolved in water after delivery. A minimum water content, compatible with not too great a tendency to deposit crystals, exists when the sucrose is about one-half inverted. Where invert sugar is ac-

crystals in a single stage vacuum pan. Usually three successive crops of crystals are grown, cooled, and centrifuged. The final mother liquor, which is resistant to further crystallization, is called blackstrap molasses. It is used principally as a feed for cattle and poultry, although substantial amounts are still fermented to produce industrial alcohol and rum.

Raw cane sugar refining. The refining of raw sugar begins with dissolution of the molasses, which exists in a thin film on the sucrose crystals in spite of the centrifugation. This step, called affination, brings the purity from about 98 to about 99%. The crystals are dissolved in hot water and percolated through bone char columns to remove color by adsorption. The bone char is washed, dried and burned to remove impurities and reused until it wears out mechanically and is discarded as fines. Even these have value as fertilizer because of their high content of calcium phosphate.

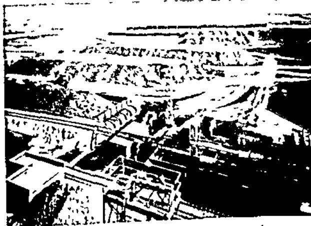
Sucrose is concentrated by vacuum evaporation, crystallized by seeding, centrifuged, and dried. In some plants, activated vegetable charcoal is used in place of bone char. The fragility of this material requires the employment of filter presses rather than char columns Recently, an activated char made from coal has become available that is sufficiently strong to be used in columns Ion exchange resins reduce the ash content of sugar solutions and thus increase sucrose recovery with a lowering of molasses production. They are employed to a limited but increasing extent.

Beet sugar. In the United States, sugar beets are grown under contract by farmers from seed supplied by a beet sugar company (*see* SUGAR BEET). Because sugar beets, like other temperate zone crops, thrive best under crop rotation, they [illegible]

to

the

slices extracted countercurrently with hot water to remove the sucrose. The resulting solution is purified by repeatedly precipitating calcium carbonate, calcium sulfite, or both in it Colloidal impurities are entangled in the growing crystals of precipitate and removed by continuous filtration The resulting solution is nearly colorless and the sucrose is concentrated by multiple-effect vacuum evaporation. The syrup is seeded, cooled, centrifuged, and the crystals are washed with water and dried. Beet molasses differs from cane molasses in having a much lower content of invert sugar. It is, therefore, relatively stable to the action of alkali and in the United States is usually treated with calcium oxide to yield a precipitate of calcium sucrate This is a mixture of loose chemical aggregates of sucrose and calcium oxide which are relatively insoluble in water. The precipitate is filtered, washed, and added to the incoming crude sugar syrup It furnishes calcium for the precipitations of calcium carbonate and sulfite, referred to above, which remove impurities. Carbon dioxide in the form of flue gas is the other reagent, and for the sulfitation step, sulfur dioxide from burning sulfur is used.



Outside a modern beet sugar factory, the beets are stored in piles to await processing *(Sugar Research Foundation, Inc.)*

Beet residue use. The beet tops and extracted slices as well as the molasses are valuable as feeds. More feed for cattle and other ruminants can be produced per acre-year from beets than from any other crop widely grown in the United States. This is independent of the food energy in the crystallized sucrose, which exceeds that available from any other temperate-zone plant. It is for these reasons that the densely populated countries of Europe have expanded their beet sugar production, in spite of the ready availability of cane sugar from the tropics.

The increased use of nitrogenous fertilizer has resulted in augmenting the protein content of the molasses and other beet by-products. In 1958 the commercial ammoniation of extracted dried beet slices was begun; this process increases still further the protein equivalent for ruminants.

Nutritional value of sugar. Sucrose has, in the past, been attacked by some nutritionists on the ground that it provides only "empty calories," without protein, minerals, or vitamins. This argument lost much of its force when it was shown that all the vitamins and minerals recommended by the National Research Council can be obtained by consuming any of a great variety of foods in amounts which yield a total of only one-half of the caloric requirements of an average person The wide use by the public of vitamin supplements has caused some nutritionists to express an opposite worry over excessive vitamin consumption The problem of sugars and tooth decay is not so easily disposed of, however, and continues to be the subject of active research. The discovery that sugars have an effect in diminishing appetite has led to their inclusion in reducing diets.

Use in organic synthesis. Since 1952 active research has been underway on the use of sucrose as an inexpensive, pure, and readily available starting material for organic synthesis. This has led to the commercial production of sucrose-based detergents, plastics, and plasticizers

Other sources of sugar. In Hawaii, the pineapple industry recovers both sucrose and citric acid from the rinds. Th [illegible] is fed t [illegible]

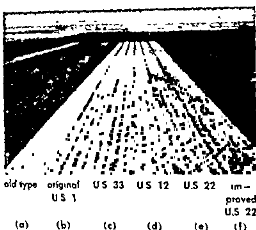


Fig 2 Demonstration of curly top resistant varieties of sugar beets in Idaho (a) The susceptible European variety, Old Type, (b) the first curly top resistant variety, U.S. 1, (c-f) various improvements up to the second release of U.S 22 (From A G Norman, ed., *Advances in Agronomy, vol. 7, Academic Press, 1955*)

of curly top in sugar beets Control of this disease, once a threat to the western beet sugar industry, has come from introduction of curly-top-resistant varieties by the U S Department of Agriculture (Fig. 2)

Leaf spot This disease, caused by the fungus *Cercospora beticola*, may be serious in Colorado and Nebraska if rainfall throughout the growing season is above normal The parasite is seed-borne; early spring infections develop on scattered seedlings If rainy periods are frequent, spores of the fungus from these primary foci spread and bring about new infections which, in turn, produce new crops of spores Thus a leaf spot epidemic may develop by mid-July from repeated sporulations and infections Blighted foliage of the sugar beet is replaced by new leaf growth that is made at the expense of root growth Sugar storage in the roots is reduced Therefore, susceptible varieties produce low yields of low quality roots that are unprofitable for both the farmer and the factory. Leaf spot resistant varieties bred by the USDA control the disease

Black root This is a general name given to the seedling diseases that bring about serious losses in humid areas Several soil-inhabiting fungi are responsible. One group known also as the cause of damping-off (stem rot near soil surface) of vegetable crops, brings about, on poorly drained fields, an acute disease of sugar beet seedlings whenever the spring season is wet Treatment of seed with a fungicide and improvement of soil conditions usually control damping-off The chronic black root caused by the water mold, *Aphanomyces cochlioides* is much more difficult to control This fungus not only causes damping-off, which may reduce the stand but also attacks the feeding roots of the sugar beet throughout the season, dwarfing affected plants Seasonal conditions and degree of soil infestation determine extent of loss from *Aphanomyces*. Crops such as legumes increase the fungus infestation, whereas corn exerts a sanitative effect. Proper crop sequences and use of black-root-resistant varieties bred by the USDA have given good control

Virus yellows. This disease is extremely important in Europe and it threatens to be equally serious in California and in the other sugar beet seed-producing districts of the United States. Various aphids are vectors of the virus (*see* APHID). Adequate control measures are not known. *See* SUGAR; SUGAR CANE. [G.H.CO]

Bibliography: See AGRICULTURAL SCIENCE (PLANT); PLANT DISEASE.

Sugar cane

Sugar cane, *Saccharum officinarum*, belongs to the grass family The present commercial canes are hybrids from the so-called "noble canes" found in the gardens of New Guinea. They probably trace their ancestry to the wild species, *Saccharum robustum*, localized in New Guinea Another wild species, *Saccharum spontaneum*, is widely distributed throughout southeast Asia.

Modern canes are the results of crosses primarily between one or more of these three species (Fig 1). The chromosome number (2N) of *S. officinarum* is 80; that of *S. robustum* varies between 60 and 148, that of *S spontaneum* varies between 48 and 128. Two other recognized species are *S. sinense* and *S Barberi*.

Structure. Every new variety of cane begins as a single shoot or culm coming from the germination



Fig 1 A field of mature, irrigated sugar cane in Hawaii, showing tasseling.

ceptable, as in the manufacture of bread, this makes for economy, a water content of only 22% in the syrup being standard.

It is the aim of the sugar refiner and the beet processor to eliminate color and all flavors other than sweetness. The manufacture of table syrups, which are widely used on waffles and pancakes, aims at a broader spectrum of flavor. Corn syrup, which is somewhat lacking in sweetness, ordinarily has about 15% sucrose added. The high viscosity of the corn syrup, resulting from the content of dextrins, tends to hinder crystallization, and is an advantage in the manufacture of certain candies.

Some sugar refiners reduce the color of their molasses and remove much of the characteristic strong flavor, thus producing an acceptable table syrup. Another source of table syrup is the juice of sorghum, which is expressed and concentrated by evaporation in open pans at atmospheric pressure. The product is dark and strongly flavored, and seems to be declining in production *See* FOOD ENGINEERING. [H.B.H.]

Bibliography: F. J. Bates, et al, *Polarimetry, Saccharimetry and the Sugars*, U.S. Nat. Bur. Standards, Circ., 440, 1942; N. Deerr, *The History of Sugar*, vols. 1 and 2, 1949–50, P. Honig (ed.), *Principles of Sugar Technology*, 1953; O. Lyle, *Technology for Sugar Refinery Workers*, 3d ed., 1957.

Sugar beet

The sugar beet, *Beta vulgaris*, originated from the white Silesian beet which Franz Carl Achard developed in 1799 from wild beets found in Sicily and along the shores of the Mediterranean. The sugar beet is an annual in southern latitudes and a biennial in the north, where it stores sugar the first year and produces seed the second (*see* ANNUAL PLANTS; BIENNIAL PLANTS). The sugar beet is an enlarged root, white in color, that gradually diminishes in diameter from crown to the tip (Fig 1) The first six inches of the tap root is free of side roots, but below that there is an extensive system of laterals. *See* ROOT (BOTANY). It has a short stem from which brilliant green, smooth leaves appear, the oldest leaves on the outside. *See* LEAF (BOTANY); STEM (BOTANY). Each leaf has prominent veins and long petioles which broaden out at the base. The inflorescence is large and branched. *See* FLOWER (BOTANY). The seed ball is a glomerule containing from one to many true seeds. *See* SEED (BOTANY).



Fig. 1. A typical sugar beet (USDA)

Sugar beet seed is planted in rows and fertilized similarly to other cultivated crops Harvesting takes place at 4–5 months of age by pulling the beets and cutting off the tops (*see* AGRICULTURAL MACHINERY) The juice contains about 14% sugar. Sugar beets are produced mostly in the temperate zone of the Northern Hemisphere.

World beet sugar production, 1956

Western Europe			7,154,000 tons
	France	21 5%	
	West Germany	17 7%	
	Italy	14 5%	
	United Kingdom	12 0%	
U S S R (Europe and Asia)			4,500,000 tons
Eastern Europe			2,856,000 tons
United States and Canada			2,119,000 tons
	California	28 6%	
	Colorado	17 0%	
	Idaho	*10 1%*	
Other			566,000 tons

The 10-year average beet sugar production in the United States, 1947–1956, was 1,625,000 tons, with an average farm value of $133,684,000. [L.D.B.]

Sugar beet diseases. Major diseases affecting the sugar beet in the United States are curly top, cercospora leaf spot, and black root. Epidemics of these may occur over wide areas when climatic conditions favor their development and spread Other diseases, such as rusts, mildews, root rots, bacterial and virus infections, although serious, are more localized *See* BACTERIA, FUNGI; PLANT VIRUS

Curly top. This virus disease occurs chiefly in sugar-beet-growing districts west of the Rocky Mountains. Affected plants are dwarfed and have curled, upturned leaves Susceptible varieties may fail almost completely because of curly top. The virus is transmitted by the beet leaf hopper, *Circulifer tenellus* (*see* LEAF HOPPER). This insect overwinters on weeds that have invaded western range lands, such as Russian thistle, various mustards, alfilaria, and *Halogeton* Many of these weeds harbor the curly top virus When range plants dry in the spring, virus-carrying insects move to irrigated fields, thereby starting the cycle



Fig 4 Effect of ratoon stunting disease on sugar cane (a) Healthy plants (b) Diseased plants.

The 10-year average cane-sugar production in the United States, 1947–1956, including Hawaii and Puerto Rico, was about 2,715,000 tons with an annual value of approximately $327,000,000

[L D B]

Sugar cane diseases. Sugar cane is subject to over 60 diseases, of which 9 have caused, or are causing, serious losses in susceptible varieties. Diseases become of particular importance in sugar cane because of (1) the use of the stalk (vegetative part) for commercial plantings, a practice that spreads disease into new plantings if the propagative stalks are diseased, (2) the relatively small number of varieties grown ... poses la... ety is st... eral crops from one original planting, which may result in the accumulation of disease in a field. Fortunately, the distribution of each of the major diseases is not world-wide, and control of all the diseases in any one country is not necessary If not brought under control, a major disease can cause losses as high as 50%.

The major diseases of sugar cane are caused by viruses, bacteria, and fungi (see BACTERIA, FUNGI, PLANT VIRUS) Viruses cause mosaic, ratoon stunting disease (Fig 4), and Fiji disease Gummosis and scald caused by bacteria, as well as red rot, smut, downy mildew, and root rot caused by fungi, are important diseases of sugar cane Except for root rot, all diseases are carried in the seed piece (stem node), and are spread in the field by insect vectors, by mechanical means such as harvesting, or by rain

Resistant varieties provide the most satisfactory means for controlling disease. Such varieties are available for all important diseases except ratoon stunting. At present ratoon stunting disease is controlled by immersing cane in hot water at 50°C for 3 hours before planting, or by hot air treatment at 58–59°C for 8 hours. *See* SUGAR; SUGAR BEET.

[S J.P C]

Bibliography: *See* AGRICULTURAL SCIENCE (PLANT); PLANT DISEASE.

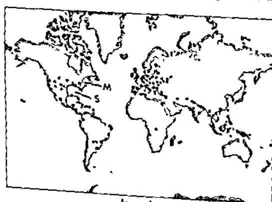
Sugar crops

All green crops produce sugar by photosynthesis. However, only a small number of plants store sufficient sugar to be of commercial importance (*see* PHOTOSYNTHESIS; PLANT METABOLISM). Sugar cane and sugar beets are the world's chief sugar crops, although sweet sorghums, sugar maples, and sugar palms have some economic significance.

Sugar cane and the sugar palms are tropical plants, the former being almost completely restricted by the 30° latitude line; sweet sorghums are confined primarily to southern United States; sugar beets are a temperate zone crop; maple sugar is found principally in northeastern United States and adjoining Canadian provinces

Palm sugar production is a village and forest industry in the provinces of Bengal and Madras in India The wild date and the palmyra, sago, and coconut palms constitute the most important species. Incisions are made in the trunk and each tree is tapped as often as 40 times over a period of 4 months, yielding about 40 lb of sugar per tree

The maple sugar and syrup industry extends from Indiana to Nova Scotia, Quebec and Ontario Provinces in Canada and New York and Vermont in the United States have the largest production. The hard maple, *Acer saccharum*, accounts for 75% of the sugar and syrup, the soft and red maples providing the other 25% of the production.



World distribution of sugar crops, exclusive of sugar palms (From H M Leppard, ed, Goode's Series of Base Maps, Univ Chicago Press, 1939)

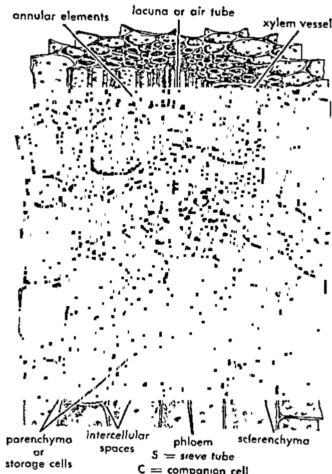


Fig. 2. The internal structure of a sugar cane internode. (*After J. P. Martin*)

World cane sugar production, 1956 (short tons)

North America, including Central America, the Caribbean area and Hawaii			12,573,000
	Cuba	6,250,000 tons	
	Hawaii	1,150,000 tons	
	Mexico	1,100,000 tons	
	Puerto Rico	1,100,000 tons	
Asia, including Philippines and Indonesia			6,104,000
	India	2,577,000 tons	
	Philippines	1,210,000 tons	
South America			5,329,000
	Brazil	2,697,000	
Africa, including adjoining islands			2,425,000
Australia and Fiji			1,503,000

of a cane seed. *See* Seed (botany). The stalk is cylindrical, is divided into nodes and internodes, and has lateral buds and an apical bud *See* Bud (botany); Stem (botany) At each node is found a leaf sheath attached to the stalk, alternately on opposite sides, and a root band which includes several rows of root primordia and a bud located alternately on opposite sides *See* Root (botany). The internodes (Fig. 2) are covered with wax and are filled with parenchyma or storage cells (pith) and the vascular bundles (*see* Parenchyma; Pith; Vascular bundles). The leaf consists of a sheath, auricle, ligule, dewlap, and a blade which tapers gradually from the base to the tip and is supported by a midrib extending almost its entire length. The edges of the leaves of most varieties are serrate. *See* Leaf (botany) The inflorescence is a silky panicle bearing many small spikelets which are arranged in pairs on the branches. *See* Flower (botany). Each spikelet contains a bisexual flower with three anthers and a single ovary surmounted by two plumelike stigmas The flower primordium is differentiated under the influence of short days or long nights (*see* Photoperiodism in plants).

Reproduction. Cane is propagated asexually with cuttings giving rise to what is known as a stool of cane (Fig. 3) consisting of roots and both secondary and primary stalks (*see* Reproduction, plant). In most countries, the crop is started during the fall, winter, or spring months, and harvested by cutting the stalks at the surface of the ground at 11–16 months of age. In Hawaii the crop periods range from about 24 months at lower elevations to 30 months or more at elevations above 2000 ft. After harvest the stool or stubble develops new shoots and shoot roots and produces a ratoon crop. A crop cycle generally includes one planted crop followed by 1–3 ratoons. The crop requires 60 in. or more of annual rainfall or irrigation, which is usually accomplished by planting the cane in furrows. It is a high consumer of nitrogen, phosphorus, and potassium. In 1956, sugar cane accounted for 61.8% of the world's sugar.

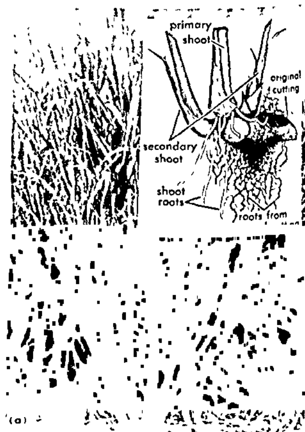


Fig 3. (a) A stool of 8-month-old sugar cane in Hawaii (b) A sugar cane cutting showing new shoots and roots.

lly Sensitization reactions, such as drug fever, rug rashes, and contact dermatitis, *are frequent.* he *incidence of skin sensitization is strongly in*rea*sed by the external* use of sulfonamide ointnents. Not rare is the development of photosensiization after internal use of sulfonamides.

Bacterial resistance. This *has developed* to all *nown sulfonamides*, and many sulfonamide re*sistant strains are encountered* among the gram-positive *and* gram-negative bacteria. In 1939, almost 100% of all cases of meningitis or gonorrhea, for instance, responded to proper sulfonamide treatment, while 20 years later, the response was less than 50%. The emergence of resistance to sulfonamides seems less rapid and less widespread than resistance against most antibiotics. *See* GONORRHEA; MENINGITIS.

Sulfonamides are used today mostly as auxiliary drugs in combination with one of the generally more effective antibiotics. In certain infectious diseases, however, for instance, in meningococcal infections and most infections of the urinary tract, *sulfonamides* deserve preference over antibiotics. Renewed *interest in* sulfonamides is due to the higher activity in the newer sulfonamides, the generally low *incidence of side effects*, and the low cost of sulfonamide medication *See* ANTIBACTERIAL AGENTS; DRUG RESISTANCE. [R.L.M.]

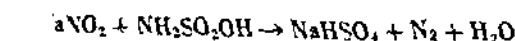
Bibliography: F Hawking and T S Lawrence *The Sulfonamides*, 1951; E. H Northey. *The Sul*

... 311–324, 1951

Sulfamate

Organic and inorganic derivatives of sulfamic acid, H_2NSO_2OH The inorganic salts which contain the $H_2NSO_2O^-$ ion are all very soluble with the exception of the basic mercury salt

The *sulfamate ion may* be determined quantitatively by *measuring* the volume of nitrogen liberated *during the* reaction



$$NaNO_2 + NH_2SO_2OH \rightarrow NaHSO_4 + N_2 + H_2O$$

Ammonium sulfamate is used to flameproof fabrics and as a *weed killer Because of* their high solubilities the sulfamates *of* nickel, copper, and lead have been used in electroplating baths

Organic sulfamates which are useful are the cyclohexylsulfamates, which are sweetening agents, and amine salts, which are used as *softeners for* paper and textiles. See SULFAMIC ACIDS [E E.WR.]

Sulfamic acids

Organosulfur compounds $RNHSO_3H$, that are organic derivatives of sulfamic acid, H_2NSO_3H Many aliphatic examples are known, but free aromatic sulfamic acids, such as phenylsulfamic acid, *are not* stable, except in the form of salts A series of stable 2 thiazolylsulfamic acids has been *prepared, however*, and their stabilities rationalized Sodium 2 thiazolylsulfamate, like sodium cyclohexylsulfamate, is very sweet and has no bitter aftertaste. The cyclohexylsulfamate salts are now widely used as sweetening agents *See* ORGANOSULFUR COMPOUND; SACCHARIN. [N K.]

Sulfanilic acid

An organic compound of the aromatic type that contains both an amino group and a sulfonic acid group, and that is sometimes called *p*-aminobenzenesulfonic acid. Its formula $H_2N—C_6H_4—SO_3H$ is better written as an inner salt,



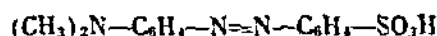
$$H_3\overset{+}{N}—C_6H_4—SO_3^-$$

Synthesis consists of heating aniline sulfate to 190°C until it rearranges. Sulfanilic acid crystallizes with one molecule of water, which it loses at 100°C. Diazotization of sulfanilic acids yields *p*-diazobenzenesulfonate



$$^-O_3S—C_6H_4—\overset{+}{N}{\equiv}N$$

the starting compound for numerous coupling *reactions* with phenols or tertiary amines to give *azo* dyes and acid base indicators For example, coupling with dimethylaniline yields methyl orange, *p,p'*-dimethylaminoazobenzenesulfonic acid,



$$(CH_3)_2N—C_6H_4—N{=}N—C_6H_4—SO_3H$$

One important use is as a starting material for the preparation of many of the sulfa drugs *See* ANILINE; ORGANOSULFUR COMPOUND; SULFA DRUGS; SULFONIC ACID [L B C.]

Sulfate

A negative ion having the formula, SO_4^{2-}, and derived from sulfuric acid, H_2SO_4. Because sulfuric *acid contains* two hydrogens, both normal sulfates and bisulfates *are known*

Of the normal sulfates, most are quite soluble *in* water, with the exception of those of silver, mercury(I), lead, strontium, barium, and calcium.

Sulfate is determined both qualitatively *and* quantitatively by precipitation as barium sulfate, $BaSO_4$, which is the most insoluble sulfate

Sodium sulfate is obtained as a by-product of the production of hydrochloric acid from salt. It is used in the manufacture of kraft paper, paperboard, and glass. *See* SULFUR; SULFURIC ACID. [E E.WR.]

Sulfenyl chlorides

A group of organosulfur compounds, RSCl, that were obscure until a *few years ago*, but are now well known They *are highly reactive*, but can generally be synthesized and isolated They undergo *numerous* reactions in which chlorine is displaced *from* sulfur by negative ions or bases, for example, cyanide ion, amines, olefins, or acetylenes. The best known ...

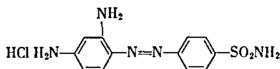
The sap contains about 3% sucrose. From 2 to 3 lb of sugar are made per tree during the season beginning in early February and ending in April. *See* SORGHUM; SUGAR; SUGAR BEET; SUGAR CANE.

[L.D.B.]

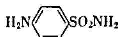
Sulfa drugs

A family of drugs of the sulfonamide type with marked ability to halt the growth of bacteria.

With the discovery of the chemotherapeutic activity of the azo dye Prontosil a new therapeutic

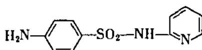


era started. For the first time it was possible to cure, by treatment with synthetic chemicals, bacterial infections which were the causes of numerous deaths, such as pneumonia, meningitis, and septicemia. The active component of Prontosil is the colorless *p*-aminobenzenesulfanilamide moiety,

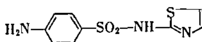


formed in the body as the result of metabolic reduction. About 50 of the many thousands of sulfonamides synthesized have been marketed, of which about 20 are still in current use.

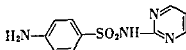
Sulfapyridine, introduced in 1937 by L. Whitby,



had highly increased antipneumococcal activity over previous sulfa drugs. Sulfathiazole, synthe-



sized in 1939 independently by various workers in America and Europe, and sulfadiazine, introduced



by M. Finland and coworkers in 1941, were superior to sulfapyridine. *See* CHEMOTHERAPY.

With sulfathiazole and sulfadiazine, the maximum of antibacterial activity was achieved, according to P. H. Bell and R. O. Roblin. They found within the group of sulfonamides a direct relationship between the electronegativity of the sulfur dioxide (SO_2) group and their bacteriostatic activity.

Sulfapyridine replaced sulfanilamide because of its considerably greater activity, and sulfathiazole replaced sulfapyridine because it caused fewer side effects, especially with respect to kidney damage. Finally, the still-better-tolerated sulfadiazine took the place of sulfathiazole, at least in the United States.

The therapeutic usefulness of a sulfonamide, however, does not depend exclusively upon its intrinsic antibacterial activity. Numerous secondary properties are important, such as solubility, especially of the metabolically formed elimination products, acetylsulfonamides, favorable absorption rates, high tissue concentration, accumulation in and penetration into affected organs, duration of circulation in the blood stream, concentration in the urine, and low incidence of side effects. Sulfaguanidine, for instance, has become a useful intestinal disinfectant because of its poor intestinal absorption. Sulfisoxazole and sulfadimetine, on the other hand, are not only rapidly absorbed and readily eliminated into the urine in effective concentrations, but their acetylation products are highly soluble in the urine. Consequently, both of these sulfonamides are excellent urinary disinfectants. Sulfamethoxypyridazine is very readily absorbed, but slowly eliminated; since a prolonged and even blood level is maintained, the daily dose necessary for effective medication is considerably reduced.

Activity and mode of action. The bacterial spectrum of sulfonamides comprises a wide variety of gram-positive and gram-negative bacteria, including staphylococci, streptococci, meningococci, and gonococci, as well as the gangrene, tetanus, coli, dysentery, and cholera bacilli. They have only slight activity against *Mycobacterium tuberculosis*, while certain closely related sulfones are quite active against *M. leprae*. The relative potency of sulfonamides against the different microorganisms varies, and their action is bacteriostatic rather than bactericidal.

The in vitro and in vivo antibacterial effect is antagonized by *p*-aminobenzoic acid (PABA) and PABA-containing natural or synthetic products, such as folic acid and procaine. Accordingly, the mode of action of sulfonamides is considered to be an antimetabolite activity, dependent upon the inhibition of enzyme systems involving the essential PABA. *See* BACILLARY DYSENTERY; CHOLERA VIBRIO; GRAM'S STAIN; LEPROSY; MENINGOCOCCUS; STAPHYLOCOCCUS; STREPTOCOCCUS; TETANUS; TUBERCULOSIS.

Side effects. The slightly soluble acetylated elimination products of the earlier introduced sulfonamides, such as sulfapyridine or sulfathiazole, were eliminated in supersaturated solution; in a number of instances, massive crystallization occurred causing obstruction of the urinary tract, particularly the kidneys, and uremic death. These accidents are very rare with the newer sulfonamides.

The danger of massive crystallization in the urinary tract can be reduced by the simultaneous administration of three different sulfonamides, a widely applied with triple sulfa preparations. The acetylated metabolites of the newer preparations, such as sulfisoxazole and sulfadimetine, are adequately soluble as such in the urine. Cyanosis, frequent with sulfanilamide, is today only rarely observed; agranulocytosis still occurs occasion

importance to the economic geologist. This system has been studied by more researchers and in more detail than any of the other binary sulfide systems. The phase relations in the Fe-S binary system are shown in Fig. 1. It is noted that vapor coexists with all phases or phase assemblages. Therefore, this diagram (Fig. 1) does not include temperature-pressure regions in the Fe-S system where vapor is not present. Since a vapor phase sometimes may be lacking during ore deposition, it is important that the temperature-pressure regions where vapor does not occur also are studied experimentally. Thus, a *complete investigation* of any sulfide system requires various experimental techniques which together include the entire geologically important temperature-pressure regions.

Experimental techniques. Because of the reactive nature of sulfur there is a very limited choice of material suitable for reaction containers within which to investigate sulfide systems.

Rigid tubes. Pure silica glass does not measurably react with sulfur, at least not below 1100°C, and has, therefore, been used extensively for sample containers in sulfide research. It should be noted that since silica is a rigid material and since the containers cannot be completely filled with the sample, a vapor space will always be present. For this reason, only the temperature-pressure regions in which vapor is a phase can be investigated by use of rigid silica reaction vessels.

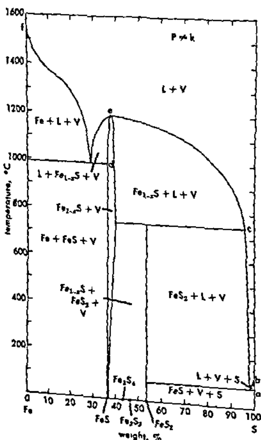


Fig. 1 Phase relations in the iron-sulfur system. All phases or phase assemblages are in equilibrium with vapor. (From G. Kullerud and H. S. Yoder, 1959)

Collapsible tubes. Sample containers, which collapse due to the application of an external pressure on the container walls, are used to investigate the temperature-pressure regions where a vapor is absent. Gold does not react with sulfur above approximately 240°C. Gold tubing collapses easily under pressure and can be readily welded, and, therefore, has been used extensively for this type of experimentation. The pressure on the sample is essentially equal to the applied external pressure because the gold container collapses. Thus, there is no free space in the container and vapor cannot form as long as the externally applied pressure exceeds the vapor pressure which would exist over the sample in a rigid tube at any given temperature.

Reaction rates and quenching procedures. The rates of reaction between metals and sulfur to form sulfides, or among the various sulfides, depend primarily upon the

upon the

being con

vacuum w

covellite, digenite, and chalcocite on copper filings in a few hours even at 25°C. However, free copper and sulfur are still present after 12 months at this temperature. The formation of copper sulfides at 100°C and under otherwise identical conditions, is much more rapid; free copper or sulfur is not visible after a few hours. At 400°C the copper sulfide formation takes place as a flash reaction.

The reactions between nickel filings and sulfur to form the Ni_3S_2, Ni_7S_6, $Ni_{1-x}S$, and Ni_3S_4 compounds are slower than those between copper and sulfur. Thus, for instance, no reaction is observed even after one month at 25°C. After three days at 100°C a film of $Ni_{1-x}S$ coats the nickel filings. At 500°C $Ni_{1-x}S$ forms, and unreacted nickel and sulfur cannot be observed after 1 hour. Formation of nickel-disulfide (NiS_2) from nickel and sulfur is a very slow process even at temperatures as high as 600°C. In such experiments $Ni_{1-x}S$ forms rapidly and is coated with NiS_2, which shields the interior of the $Ni_{1-x}S$ grains from the sulfur. Further reaction is very slow and is dependent upon the diffusion of sulfur through the disulfide shielding. On the other hand, NiS_2 will form very rapidly at temperatures above 300°C if finely ground NiS and S are mixed before heating. For example, at 600°C NiS_2 forms and unreacted NiS or sulfur cannot be observed after 2 min. Very similar observations have been made for the formation of the monosulfide and disulfide of iron. FeS or $Fe_{1-x}S$ at 500°C forms in a day from iron and sulfur. At 600°C FeS_2 forms in less than 2 min from mixtures of finely ground FeS and sulfur, but requires several weeks to form from iron and sulfur at the same temperature. (Copper,

action of sulfenyl chlorides with amines gives sulfenamides, some of which are useful as rubber vulcanization accelerators. *See* MERCAPTAN; ORGANOSULFUR COMPOUND. [N.K.]

Bibliography: N. Kharasch, The unique properties of 2,4-dinitrobenzenesulfenyl chloride, *J. Chem. Ed.*, 33:585–591, 1956.

Sulfide

A negative ion having the formula S^{2-}, and derived from hydrogen sulfide or hydrosulfuric acid, H_2S.

Because hydrosulfuric is a dibasic acid, sulfides, M_2S, and bisulfides, MHS, are formed. Although most of the metal sulfides are insoluble in water, they dissolve in acids with the evolution of H_2S gas. This escaping gas will cause paper moistened with a solution of lead acetate to turn black, a test that can be used for the presence of sulfide ion in the original solution.

The insoluble metal sulfides are used to separate and identify metal ions in qualitative analysis. Soluble sulfide solutions are very basic because of the tendency of the sulfide ions to accept a proton from water as follows:



$$H_2O + S^{2-} \rightleftharpoons HS^- + OH^-$$

See SULFUR. [E.E.WR.]

Sulfide phase equilibria

Sulfide ore deposits are the most important sources of numerous metals such as lead, zinc, copper, nickel, cobalt, and molybdenum. In addition, sulfide ores also provide substantial amounts of noble metals such as platinum, gold, and silver, and of other industrially important elements such as cadmium, rhenium, and selenium. Although iron sulfides usually are the most common minerals in such deposits, most commercial iron is mined from iron oxide ores and most sulfur from elemental sulfur deposits.

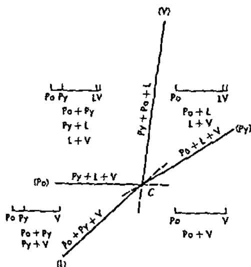
Mineral name	*Chemical formula*
Pyrite and marcasite	FeS_2
Pyrrhotite	$Fe_{1-x}S$
Covellite	CuS
Digenite	Cu_9S_5
Chalcocite	Cu_2S
Bornite	Cu_5FeS_4
Chalcopyrite	$CuFeS_2$
Cubanite	$CuFe_2S_3$
Galena	PbS
Sphalerite and wurtzite	ZnS
Metacinnabar and cinnabar	HgS
Argentite and acanthite	Ag_2S
Molybdenite	MoS_2
Millerite	$Ni_{1-x}S$
Vaesite	NiS_2
Pentlandite	$(Fe,Ni)_9S_8$

Some of the more common sulfide minerals are listed above, together with their chemical formulas. It is noted that the chemistry of the sulfides is rather simple inasmuch as most minerals involve only two major elements; some involve three and a few four. Understanding of the phase relations between these minerals is important both to the geologist whose task it is to locate and exploit ore deposits and to the metallurgist whose task it is to extract the metals from the ores for industrial use. In the following discussion sulfide phase equilibria will be dealt with from a geological point of view.

It should be noted in the list of sulfide minerals that the eight sulfides first listed may be plotted in the ternary system copper-iron-sulfur. Similarly the first two minerals and the last three may be plotted in the ternary system iron-nickel-sulfur. Thus, by studying in detail the phase equilibria in two ternary systems a great deal of information can be obtained about many of the common sulfides occurring in ore deposits. However, before the ternary systems can be explored in a systematic way, the binary systems bounding the ternaries have to be studied in detail. Similarly, it is necessary that the four bounding ternary systems are fully understood before a quaternary system can be systematically investigated. Thus, it is seen that before a systematic study of, for example, the immensely important quaternary system copper-iron-nickel-sulfur (Cu-Fe-Ni-S) can proceed, much preliminary information is required. The prerequisite data include complete knowledge of the four ternary systems Cu-Fe-S, Cu-Fe-Ni, Fe-Ni-S, and Cu-Ni-S The phase relations in these ternary systems in turn cannot be systematically studied before the six binary systems Fe-Cu, Fe-Ni, Cu-Ni, Fe-S, Cu-S, and Ni-S have been thoroughly explored.

The enormous differences in the vapor pressures over the different phases occurring in sulfide systems add complications to the diagrammatic representation. For instance, in the Fe-S system the vapor pressure over pure iron is about 10^{-25} atm at 450°C, while that over pure sulfur is a little more than 1 atm at the same temperature. A correct diagrammatic representation of the relations in such a system, therefore, requires coordinates for composition and temperature, as well as for pressure. In a two-component system, such as the Fe-S system, this is feasible because only three coordinates are necessary. However, in ternary (where such diagrams involve four-dimensional space) and in multicomponent systems, such a diagrammatic representation is not possible. For this reason it is customary to use composition and temperature coordinates only for the diagrammatic representation of sulfide systems. It should be realized that the relations as shown in such diagrams in reality represent a projection from composition-temperature-pressure space onto a two-dimensional composition-temperature plane or onto a three-dimensional prism, depending upon whether the system contains two or three components

Pyrite or pyrrhotite, or both, occur ubiquitously not only in ore deposits but in nearly all kinds of rocks as well Of the binary systems mentioned above, therefore, the iron-sulfur system is of most



ig 2 Univariant curves and divariant regions about nvariant point c where pyrrhotite, pyrite, liquid and apor are in equilibrium (From G Kullerud and H. S. Yoder, 1959)

curve of pyrite The pressure at the incongruent melting point of pyrite (743°C) can be obtained by extrapolation of any two of the curves shown in Fig 2 to their point of intersection, the invariant point.

The pyrite ⇄ pyrrhotite + liquid univariant curve (Fig 3), as well as parts of the pyrite ⇄ pyrrhotite + vapor univariant curve, has been determined The latter curve has been studied from 575°C where the vapor pressure is 2 mm Hg to 690°C where the pressure is 760 mm Hg The pressure at the point of intersection (743°C) of the two univariant curves was by extrapolation found to be about 10 bars

Very similar relations are encountered in the copper-sulfur system by the *breakdown of covellite* to digenite + vapor, or, when vapor is absent, to digenite and liquid (or gas), as well as by the decomposition of digenite to chalcocite and vapor, or, when vapor is absent, to chalcocite and liquid (or gas) Incongruent melting of covellite takes place at 507°C and at 925 mm Hg At this invariant point the four phases, digenite, covellite, liquid, and vapor, are stable. Of the four univariant curves originating in this invariant point only two have been studied The digenite, covellite, liquid (or gas) univariant curve (the upper stability curve of covellite) has been found to go through 510°C at 7500 psi, 515°C at 15 000 psi, and 525°C at 30,000 psi The digenite, covellite, vapor univariant curve has been determined from 400°C where the vapor pressure is 1 5 mm Hg to 490°C where the vapor pressure is 510 mm Hg Similarly at the incongruent melting point of digenite, invariancy is represented by the four phases chalcocite, digenite, liquid and vapor None of the four univariant curves originating in this invariant point has been determined.

Ternary systems. The copper-iron-sulfur system is bounded on two sides by the two binary systems discussed above. It is not only theoretically and economically very important, but probably is also the most complicated of the ternary systems containing sulfides of common occurrence in ore deposits. Parts of the system have been studied by H. E. Merwin and R. H. Lombard in 1937; H. Schlegel and A. Schüller in 1952; and by J. W. Greig, E. Jensen, and Merwin in 1955. In the first of these studies digenite, Cu_9S_5, was not reported as a phase (The mineral digenite strongly resembles chalcocite in polished sections and was therefore not recognized as a phase by Merwin and Lombard.) The second investigation did not cover the digenite part of the system, and the third study took place at temperatures where digenite is not stable The phase diagram presented by Merwin and Lombard is shown in Fig. 4. Since digenite does not appear as a phase in this diagram, the phase relations as shown in Fig. 4 are not entirely correct.

The ternary phases are Cu_5FeS_4 (bornite), Cu_5FeS_6 (idaite), $CuFeS_2$ (chalcopyrite), $Cu_3Fe_4S_6$, and $CuFe_2S_3$ (cubanite). The compound $Cu_3Fe_4S_6$ might be synonymous with the mineral valleriite.

In Fig. 4 the areas containing one condensed phase are marked with heavy vertical lines, except

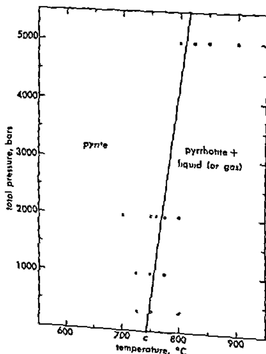


Fig 3 The upper stability curve of pyrite is the univariant curve for the reaction $FeS_2 \rightleftharpoons Fe_{1-x}S + L$. The curve originates at point c, 743°C and about 10 bars (From G Kullerud and H. S. Yoder, 1959)

nickel, and iron filings of closely the same size were used to obtain comparable reaction rates.)

The silica tubes are heated in horizontal or vertical furnaces controlled within ±1°C and wound in such a way that the temperature gradient over the length of the silica tubes (commonly about 1.5 in.) does not exceed 2°C even at 900°C. The tubes are quenched by being dropped in cold water and thus will reach room temperature in 3–5 sec. The vapor pressure in the rigid silica tubes falls during the quench at a rate dependent on the temperature and free space in the tube. Even at such a rapid rate it has not been possible to prevent many reactions, such as exsolutions and polymorphic transformations, from taking place in certain sulfide systems during the quenching period.

The gold tubes are heated in cold-seal pressure vessels, in Tuttle-type bombs, or in internally heated pressure vessels. In such experiments the pressure on the sample is fixed by applying a known external pressure on the gold tube wall by means of a gas such as argon or a liquid such as H_2O The collapsible tubes in the cold-seal pressure vessels are quenched from 800°C in about 2 or 3 min by spraying the vessels with compressed air until the temperature falls below about 600°C and then with water until cold. The collapsible tubes in the Tuttle-type pressure vessels are quenched from 800°C in about 45 sec by spraying the vessels with compressed air and then water as described above. The collapsible tubes in the internally heated pressure vessel are quenched by turning off the furnace power The pressure vessel is continuously cooled by water, so that from a run temperature of 800°C, room temperature is reached in about 30 sec after the furnace power is cut off.

The pressures in the cold-seal, Tuttle, and internally heated pressure vessels are maintained constant during the quench. The rate of quenching of the sample in the various types of pressure vessels is much slower than the rate of quenching of the silica tubes, which are dropped swiftly into cold water. Because of these slower cooling rates quenching reactions are more commonly observed in the sulfides investigated by using pressure equipment than in the rapidly quenched sulfides synthesized in silica tubes

Identification of phases. The sulfides synthesized in silica tubes, as well as those produced in collapsible gold tubes, are usually identified at room temperature by various properties such as macroscopic colors, magnetic susceptibilities, and hardness, by optical properties in plain and polarized reflected light (polished sections), and by x-ray powder diffraction patterns. No one single method of those mentioned above is adequate to definitely identify all phases which may occur or to explain the relations between the phases. It is, therefore, imperative that all products from all experiments be investigated by means of x-ray powder diffraction patterns as well as by means of polished sections. See X-RAY POWDER METHODS.

Phase relations in some sulfide systems. In a short review of this kind, it is impossible to deal with all the geologically important sulfide systems. For this reason the discussion will be confined to a few pertinent binary and ternary systems

Binary systems. The phase relations in the Fe-S system are shown in Fig 1. (Fe_2S_3, Fe_3S_4 and Fe_9S_8, which are all suggested as phases in the literature, have been omitted from Fig. 1 because their stability fields, if any, are not known) The coexistence of vapor with all phases, or phase assemblages, demonstrates that the experiments were all conducted in rigid tubes where vapor was always present. Thus the pressure-temperature regions where vapor is absent are lacking. Figure 1 shows that pyrite melts incongruently at 743°C (c) At precisely that temperature four phases are in equilibrium: pyrrhotite, pyrite, liquid, and vapor According to the Phase Rule, four coexisting phases in a two-component system constitute an invariant condition, a condition satisfied at only one pressure and one temperature. The arguments concerning invariancy, univariancy, etc., are of vital importance to an understanding of sulfide systems in general and will therefore be discussed in some detail The pressure or temperature can change from the invariant point only if one or more phases are consumed, and the system then becomes univariant or multivariant, respectively. Four univariant curves will originate from the invariant point, one for each possible three-phase assemblage

The stable assemblages along the univariant curves are as follows, the absent phase in each case being enclosed in parentheses:

(Liquid) pyrite ⇄ pyrrhotite + vapor
(Vapor) pyrite ⇄ pyrrhotite + liquid
(Pyrite) pyrrhotite + vapor ⇄ liquid
(Pyrrhotite) pyrite + vapor ⇄ liquid

The sequence of the curves about the invariant point may be deduced from the principles of G W Morey and E D Williamson and the Morey Schreinemaker's coincidence theorem. The slopes of the curves may be estimated from general considerations of the entropy and volume relations of the various phases. The curves so derived are shown in Fig 2. The two-phase assemblages in the divariant regions between the univariant curves are deduced by considering the permissible phases which will account for all possible compositions in the two component system The "composition bar" inset in the divariant regions of Fig 2 is a convenient device for recording the possible assemblages for the complete range of bulk composition

The univariant curve along which vapor is absent, pyrite ⇄ pyrrhotite + liquid, is shown in Fig 3. (The reaction above the critical pressure of sulfur (118 ± 31 bars) is by definition FeS_2 ⇄ $Fe_{1-x}S$ + gas.) It was determined by experiment in collapsible gold tubes. The reaction temperature rises at the rate of about 14°C/1000 bars (about 7 bars/°C). This curve is called the upper stability

emperatures, the solubility of FeS in ZnS is appreciable

The FeS-ZnS system is shown in Fig 5. The solvus curve GL relates the composition of the (Fe,Zn)S mix-crystals when formed in equilibrium with FeS to the temperatures which existed when the mix-crystals were formed. Similarly the composition of (Fe,Zn)S mix-crystals formed in equilibrium with FeS in ore bodies does, by application of the GL solvus curve, indicate the temperature at which the crystals were deposited, provided no change took place in the mix-crystals after their formation. Since pyrrhotite in nature is iron deficient ($Fe_{1-x}S$, not FeS), the pyrrhotite-sphalerite join is not strictly binary However, the FeS-ZnS solvus and the $Fe_{1-x}S$-ZnS solvus are found by experiments to coincide within the limits of error of

...ture and secondarily of pressure Under certain conditions such mix-crystals may serve as reliable geological thermometers For a discussion of solid-solution sulfide thermometers, as well as the many systems under exploration and the conditions under which such phase relations may be applied to natural deposits, see GEOLOGICAL THERMOMETRY.

[G K.]

Sulfinic acid

One of a group of organosulfur compounds, RSO_2H, that possess one less oxygen than sulfonic acids, and are easily oxidized to the latter In contrast to sulfonic acids, sulfinic acids undergo self-oxidation-reduction (dismutation)

$$2RSO_2H \rightarrow RSO_3H + (RSOH)$$



... (a thiolsulfonate ester) Because of instabilities, sulfinic acids are often prepared and used as salts such as sodium p-toluenesulfinate

Aromatic sulfinic acids are better known and more stable than those of the aliphatic series. Only one free aliphatic sulfinic acid, 1-dodecane...

... electroplating baths, as redox polymerization catalysts, and as reducing agents Rongalite, $HOCH_2SO_2Na$, is an important commercial reducing agent See ORGANOSULFUR COMPOUND; POLYMERIZATION

[N K.]

Sulfite

A negative ion having the formula SO_3^{2-}, and derived from the unstable sulfurous acid, H_2SO_3

Because the sulfur in the sulfite ion has a 4+ oxidation state, it will undergo oxidation to the sulfate ion or reduction to free sulfur or to the sulfide ion

The sulfite ion can be identified by the fact that SO_2 gas is liberated when solutions are acidified.



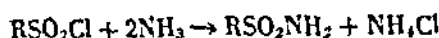
$$2H^+ + SO_3^{2-} \rightleftharpoons [H_2SO_3] \rightleftharpoons H_2O + SO_2$$

Because sulfurous acid is dibasic, two series of salts, the normal M_2SO_3 and acid $MHSO_3$, are formed.

Most of these salts, with the exception of the alkali metal and ammonium salts, are only slightly soluble. Sodium sulfite is used in dyeing and bleaching and as a preservative, other sulfites are used extensively in the manufacture of certain types of paper. See SULFUR [E.E.WR]

Sulfonamide

One of a group of organosulfur compounds, RSO_2NH_2, that are readily prepared by the reaction of sulfonyl chlorides and ammonia



$$RSO_2Cl + 2NH_3 \rightarrow RSO_2NH_2 + NH_4Cl$$

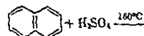
There is great interest in these substances because of the therapeutic sulfa drugs, but they are also of chemical value The Hinsberg procedure of organic analysis converts amines to sulfonamides with p-toluenesulfonyl chloride as the reagent. A primary amine is distinguishable from a secondary amine because the resultant sulfonamide ($RNHSO_2$-p-tolyl) from the primary amine still has an acidic H atom attached to nitrogen, and hence, is soluble in aqueous alkali, whereas $R_2N{\cdot}SO_2$-p-tolyl is not soluble. See AMINE; ORGANOSULFUR COMPOUND; SULFA DRUGS; SULFONYL CHLORIDE.

[N.K.]

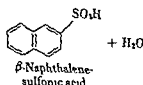
Sulfonation

A chemical reaction in which a sulfonic acid group ($-SO_3H$) is introduced into the structure of a molecule or ion in place of a hydrogen atom

Sulfonation of aromatic compounds is the most important type of sulfonation. This is accomplished by treating the aromatic compound with sulfuric acid, usually containing sulfur trioxide (solutions of SO_3 in sulfuric acid are called oleum, or fuming sulfuric acid)



Naphthalene



β-Naphthalene-sulfonic acid

This is an *electrophilic* aromatic substitution reaction, the effective electrophilic reagent is believed to be the SO_3 molecule. The product of sulfonation is a sulfonic acid.

[J T.B.]

INDUSTRIAL ASPECTS

Sulfonation may also be defined as any chemical process by which the sulfonic acid group

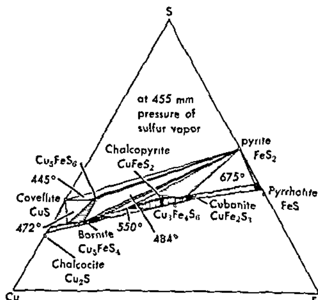


Fig. 4. Phase relations in the copper-iron-sulfur system. *(From H. E. Merwin and R. H. Lombard, The system, Cu-Fe-S, Econ. Geol. vol. 32, 2d suppl., 1937)*

the Cu_5FeS_6 area, which is marked by a horizontal line extending a short distance toward CuS At the pressure and temperatures for which the diagram is drawn, chalcocite and bornite form a complete solid solution series. The intermediate field shows solid solution from cubanite to a point almost half way between chalcopyrite and bornite. Chalcopyrite is commonly regarded as $CuFeS_2$, but is in reality a sulfur-deficient compound For this reason $CuFeS_2$ is plotted outside the intermediate solid solution field. The third field containing only one condensed phase is the pyrrhotite field, which appears to the right in Fig 4. Of the fields containing two condensed phases, eight are marked with horizontal lines and the CuS-Cu_5FeS_6 and CuS-FeS_2 fields are shown by tie lines. The five clear areas in the diagram, labeled with temperatures, represent fields where three condensed phases are present. All phases or phase assemblages in Fig. 4 are in equilibrium with vapor. Thus, the three condensed phases in clear areas, which are in equilibrium with vapor, form univariant assemblages

The Fe-Ni-S system, bounded on two sides by the Fe-S and Ni-S systems, is also of great theoretical and economic importance. This system has been studied in much less detail than the Cu-Fe-S system The minerals involved in this system are pyrite and pyrrhotite on the Fe-S side; vaesite (NiS_2), polydymite (Ni_3S_4), millerite ($Ni_{1-x}S$), and heazlewoodite (Ni_3S_2) on the Ni-S side; as well as the ternary compounds violarite (Ni_2FeS_4) and pentlandite $(Fe,Ni)_9S_8$. Most of the world's production of nickel comes from pentlandite, a mineral which almost without exception occurs intimately intergrown with pyrrhotite in ore deposits. Systematic studies designed to gain an understanding of the pyrrhotite-pentlandite relations at various temperatures and pressures are progressing at the Geophysical Laboratory, Carnegie Institution of Washington.

In the ternary system Fe-Zn-S, the minerals pyrrhotite and pyrite occur along the Fe-S boundary and the mineral sphalerite (ZnS) occurs along the Zn-S boundary. (ZnS has at least two polymorphs sphalerite being stable below 1020°C and wurtzite being stable above 1020°C.) This system is important because the minerals sphalerite, pyrrhotite and pyrite are of common occurrence in sulfide ore deposits and because of the solid solutions these minerals form. Thus in numerous ore deposits sphalerite and pyrrhotite occur together, many carry sphalerite and pyrite, and in some deposits sphalerite and pyrrhotite as well as pyrite are found in coexistence. Although no ternary compounds, structurally different from those occurring along the binary joins, are known in the system, the studies of the solid solutions which extend appreciably into the ternary field have been very rewarding. Thus the results of a systematic experimental study of the extent of solid solution between FeS and ZnS at various temperatures and pressures have been used extensively in geological thermometry. These experiments showed that although the amount of ZnS which will dissolve in FeS is insignificant at all geologically important

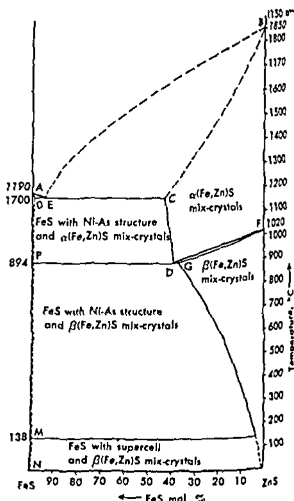
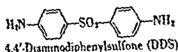


Fig 5. The FeS-ZnS equilibrium diagram. All phases or phase assemblages are in equilibrium with vapor *(From G. Kullerud, The FeS-ZnS system; a geological thermometer. Norsk Geol. T. 32.61–1647, 1953*

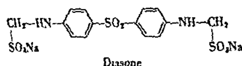
lfone (antimicrobial use)

s group of chemical compounds is used in the atment of leprosy and tuberculosis. In 1937, in- tigators in England and France showed inde- idently that 4,4'-diaminodiphenylsulfone, or)S, was very effective in curing *streptococcal* in- tions in mice Although this compound was st synthesized in 1908, it was examined as a pos- le chemotherapeutic agent only when it was und as an impurity in the manufacture of sulfa- lamide While DDS is 100 times more potent than

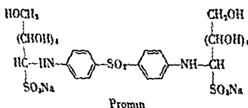


4,4'-Diaminodiphenylsulfone (DDS)

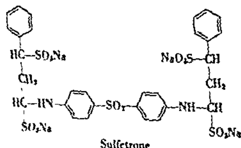
ulfanilamide against *Streptococcus pyogenes* in ice, it is considerably more toxic; and when it as first used in man for the treatment of acute treptococcal infection, it had to be discontinued ecause of the severe hemolytic anemia which de- eloped Attempts were made to produce less toxic derivatives, and a number of analogs with substitu- ents on the amino groups were synthesized



Diasone



Promin



Sulfetrone

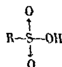
In 1940, DDS was found to be active against tuberculosis in rabbits and promin was shown to be effective in guinea pigs infected with human tubercle bacilli In man, however, promin had disappointing antitubercular activity It was probably the reports of its antitubercular activity in animals that prompted its trial in 1941 in the treatment of leprosy. This marked the beginning of the most significant advance yet made in the treatment of leprosy. DDS is now the most generally used sulfone in leprosy and is the most effective treatment for this disease. There is strong evidence to show that, in the body, all the disubstituted derivatives of DDS mentioned above are converted to the parent substance See CHEMOTHERAPY; LEPROSY; TUBERCULOSIS. [N.J.C.]

Sulfone (chemical)

One of a group of organosulfur compounds, RSO_2R', that are well known, stable, and generally crystalline substances. Sulfones are best prepared by the oxidation of sulfides, and also by the reaction of sulfonyl chlorides (RSO_2Cl) with aromatic compounds in the presence of Friedel-Crafts catalysts, or by reaction of sodium sulfinates and alkyl bromides Sulfones resist further oxidation, but can be reduced, under special conditions, to sulfides. In the presence of Raney nickel, with absorbed hydrogen, RSO_2R' is reduced to $RH + R'H$ through a carbon-sulfur bond-scission process Pyrolysis of sulfones frequently involves elimination of sulfur dioxide in a synthetically useful way. For example, alkenyl aryl sulfones smoothly yield alkenes. The sulfone group exerts a strong inductive effect and increases the acidity of α-hydrogen atoms. In $(CH_3SO_2)_3CH$, the acidity becomes very high. The $-SO_2-$ group also deactivates the benzene ring to attack by electrophilic reagents. Numerous practical uses have been claimed for variously substituted sulfones in recent years *See* ORGANIC REACTION MECHANISM; ORGANOSULFUR COMPOUND; SULFIDE; SULFOXIDE. [N.K.]

Sulfonic acid

One of a group of organosulfur compounds,



that are exemplified by methanesulfonic acid, CH_3SO_3H, and benzenesulfonic acid, $C_6H_5SO_3H$. They are strongly acidic, water soluble, nonvolatile, and hygroscopic, and they do not act as oxidizing agents. The aliphatic sulfonic acids are made by oxidizing mercaptans or disulfides (RSH or [illegible]
me [illegible]
anc [illegible]
as a by-product from petroleum refining Other aliphatic sulfonic acids are also known but have not been as extensively studied as the aromatic compounds 10-Camphorsulfonic acid, derived from camphor, is well known, and certain substituted [illegible]

—SO_2OH, or the corresponding salt or sulfonyl halide group (for example, —SO_2Cl) is introduced into an organic compound. These groups may be bonded to either a carbon or a nitrogen atom. The latter compounds are designated *N*-sulfonates or sulfamates.

Sulfation involves the attachment of the —OSO_2OH group to carbon, yielding an acid sulfate ($ROSO_2OH$), or of the —SO_4— group between two carbons, forming the sulfate ($ROSO_2OR$).

Uses of sulfonates and sulfates. Millions of tons of sulfonates are manufactured annually. Most sulfonates are employed as such in acid or salt form for applications where the strongly polar hydrophilic —SO_2OH group confers needed properties on a comparatively hydrophobic nonpolar organic molecule. Some sulfonates, such as methane- and toluenesulfonic acids, are used as catalysts. A relatively large number of sulfonates are marketed in salt form but used in acid form; such compounds include dyes, mothproofing agents, and synthetic tanning agents. In these cases, the salts are applied in acid medium, thereby permitting the free —SO_2OH group of the organic molecule to become attached to the textile fiber or leather. The major quantity of sulfonates and sulfates is both marketed and used in salt form. This category includes detergents, emulsifying, demulsifying, wetting, and solubilizing agents, lubricant additives, and rust inhibitors.

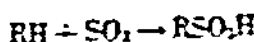
Aromatic sulfonyl chlorides (—$ROSO_2Cl$) are useful for preparing sulfonamides (including sulfa drugs, dyes, tanning agents, plasticizers, and the sweetening agent saccharin).

Sulfonates and sulfates are employed for the preparation of organic compounds devoid of sulfur. Thus phenol, resorcinol, and naphthols are obtained by the caustic fusion of the parent sulfonates, whereas ethanol and isopropanol are obtained by the hydrolysis of sulfated alkenes, for example, ethyl hydrogen sulfate.

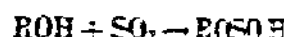
Sulfonating and sulfating agents. The principal agents are as follows: (1) sulfur trioxide and compounds: (*a*) sulfur trioxide, oleum ($H_2SO_4 + SO_3$), and concentrated sulfuric acid (SO_3 + water); (*b*) chlorosulfonic acid ($SO_3 + HCl$); (*c*) sulfur trioxide adducts with organic compounds (SO_3 + dioxane); (*d*) sulfamic acid; and (2) the sulfur dioxide group: (*a*) sulfurous acid and metal sulfites; (*b*) sulfur dioxide with chlorine or oxygen.

Sulfur trioxide, oleum, and concentrated sulfuric acid are considered together because of their close physical relationship and because they can, in certain cases, be used interchangeably. This group accounts for the preponderant production of aromatic sulfonates.

Sulfur trioxide is theoretically the most efficient sulfonating and sulfating agent, because only direct addition is involved.



$$RH + SO_3 \rightarrow RSO_3H$$



$$ROH + SO_3 \rightarrow ROSO_3H$$

However, sulfur trioxide combines with water and aromatic compounds with the evolution of so much heat that its activity must be moderated. For this reason, the hydrates of SO_3 (oleum and sulfuric acid) are generally used. Here, SO_3 is the true reactive species, the water functioning only as a complexing agent and a solvent. An increase in the water content lowers the activity of the reagent in sulfonation; the reaction rate is inversely proportional to the square of the water concentration. When the SO_3 concentration in the sulfonating agent has been reduced to a critical level which depends upon the organic compound being treated, sulfonation stops. The critical concentrations (π values) for the monosulfonation of naphthalene, benzene, and nitrobenzene are approximately [illegible] 64, and 82% SO_3.

Catalysts. The addition of certain chemicals, usually in small amounts, can have a marked influence on some sulfonations. The addition of mercury changes the orientation in a number of aromatic sulfonations. This is of great importance in the preparation of α-anthraquinone sulfonates. In the absence of mercury compounds, the β-sulfonates are obtained exclusively. About 1% mercury as metal or salt equivalent, based on anthraquinone used is required. Mercury also affects the orientation in the sulfonation of benzoic acid, phthalic anhydride, and nitrobenzene. In these reactions the quantity of mercury needed is high, and the influence on orientation is only partial.

Aromatic sulfonation, like nitration and halogenation, is a typical electrophilic substitution reaction. Sulfonation, however, differs from these other reactions in two respects: it is reversible, and in certain cases (for example, in the sulfonation of naphthalenes) temperature has an important influence on the position of the entering group.

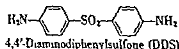
Equipment. Cast iron is resistant to the action of sulfuric acid in the range 75–100% in strength over a fairly wide temperature range and has been a standard material of construction for sulfonation kettles for many years especially for numerous dye intermediates and for aromatic hydrocarbons. However, it has poor tensile strength and is corroded by oleum or sulfur trioxide. This fault may be controlled in oleum sulfonations by adding acid slowly to the material being sulfonated to keep the acid concentration below the corrosive level.

The use of lined steel vessels combines low cost and high strength with good corrosion resistance. Commonly used linings include glass, enamel, lead, and type 316 stainless steel.

Continuous operations involving special equipment are advantageous only where the reaction rate is fast and where the volume of production (as in benzenesulfonic acid and dodecylbenzene sulfonate) is large and relatively steady. See ORGANIC SULFUR COMPOUND; SUBSTITUTION REACTION; UNIT PROCESSES. [illegible]

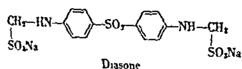
Sulfone (antimicrobial use)

This group of chemical compounds is used in the treatment of leprosy and tuberculosis. In 1937, investigators in England and France showed independently that 4,4'-diaminodiphenylsulfone, or DDS, was very effective in curing streptococcal infections in mice. Although this compound was first synthesized in 1908, it was examined as a possible chemotherapeutic agent only when it was found as an impurity in the manufacture of sulfanilamide. While DDS is 100 times more potent than

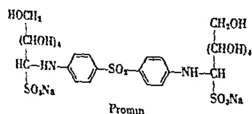


4,4'-Diaminodiphenylsulfone (DDS)

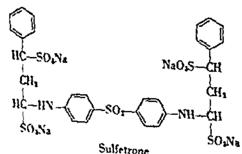
sulfanilamide against *Streptococcus pyogenes* in mice, it is considerably more toxic; and when it was first used in man for the treatment of acute streptococcal infection, it had to be discontinued because of the severe hemolytic anemia which developed. Attempts were made to produce less toxic derivatives, and a number of analogs with substituents on the amino groups were synthesized.



Diasone



Promin



Sulfetrone

In 1940, DDS was found to be active ... tuber... ... ...ercular activity. It was probably the reports of its antitubercular activity in animals that prompted its trial in 1941 in the treatment of leprosy. This marked the beginning of the most significant advance yet made in the treatment of leprosy. DDS is now the most generally used sulfone in leprosy and is the most effective treatment for this disease. There is strong evidence to show that, in the body, all the disubstituted derivatives of DDS mentioned above are converted to the parent substance. See CHEMOTHERAPY; LEPROSY; TUBERCULOSIS. [N.J.C.]

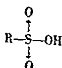
Sulfone (chemical)

One of a group of organosulfur compounds, RSO_2R' that ...

compounds in the presence of Friedel-Crafts catalysts, or by reaction of sodium sulfinates and alkyl bromides. Sulfones resist further oxidation, but can be reduced, under special conditions, to sulfides. In the presence of Raney nickel, with absorbed hydrogen, RSO_2R' is reduced to $RH + R'H$ through a carbon-sulf... sulfones freq... dioxide in a s... alkenyl aryl ... sulfone group exerts a strong inductive effect and increases the acidity of α-hydrogen atoms. In $(CH_3SO_2)_3CH$, the acidity becomes very high. The $—SO_2—$group also deactivates the benzene ring to attack by electrophilic reagents. Numerous practical uses have been claimed for variously substituted sulfones in recent years. See ORGANIC REACTION MECHANISM; ORGANOSULFUR COMPOUND; SULFIDE; SULFOXIDE. [N.K.]

Sulfonic acid

One of a group of organosulfur compounds,



that are exemplified by methanesulfonic acid, CH_3SO_3H, and benzenesulfonic acid, $C_6H_5SO_3H$. They are strongly acidic, water soluble, nonvolatile, and hygroscopic, and they do not act as oxidizing agents. The aliphatic sulfonic acids are made by oxidizing mercaptans or disulfides (RSH or RSSR). Methane sulfonic acid has been recommended for catalyzing esterifications, hydrolyses, and alkylations. The acid is available commercially as a by-product from petroleum refining. Other aliphatic sulfonic acids are also known but have not been as extensively studied as the aromatic compounds. 10-Camphorsulfonic acid, derived from camphor, is well known, and certain subst... aliphatic ... ethion ... of spe

—SO_2OH, or the corresponding salt or sulfonyl halide group (for example, —SO_2Cl) is introduced into an organic compound. These groups may be bonded to either a carbon or a nitrogen atom. The latter compounds are designated *N*-sulfonates or sulfamates

Sulfation involves the attachment of the —OSO_2OH group to carbon, yielding an acid sulfate ($ROSO_2OH$), or of the —SO_4— group between two carbons, forming the sulfate ($ROSO_2OR$).

Uses of sulfonates and sulfates. Millions of tons of sulfonates are manufactured annually. Most sulfonates are employed as such in acid or salt form for applications where the strongly polar hydrophilic —SO_2OH group confers needed properties on a comparatively hydrophobic nonpolar organic molecule. Some sulfonates, such as methane- and toluenesulfonic acids, are used as catalysts. A relatively large number of sulfonates are marketed in salt form but used in acid form, such compounds include dyes, mothproofing agents, and synthetic tanning agents. In these cases, the salts are applied in acid medium, thereby permitting the free —SO_2OH group of the organic molecule to become attached to the textile fiber or leather. The major quantity of sulfonates and sulfates is both marketed and used in salt form. This category includes detergents, emulsifying, demulsifying, wetting, and solubilizing agents, lubricant additives, and rust inhibitors.

Aromatic sulfonyl chlorides (—$ROSO_2Cl$) are useful for preparing sulfonamides (including sulfa drugs, dyes, tanning agents, plasticizers, and the sweetening agent saccharin).

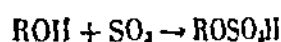
Sulfonates and sulfates are employed for the preparation of organic compounds devoid of sulfur Thus phenol, resorcinol, and naphthols are obtained by the caustic fusion of the parent sulfonates, whereas ethanol and isopropanol are obtained by the hydrolysis of sulfated alkenes, for example,

SO_3), and concentrated sulfuric acid (SO_3 + water); (*b*) chlorosulfonic acid (SO_3 + HCl); (*c*) sulfur trioxide adducts with organic compounds (SO_3 + dioxane); (*d*) sulfamic acid; and (2) the sulfur dioxide group: (*a*) sulfurous acid and metal sulfites; (*b*) sulfur dioxide with chlorine or oxygen

Sulfur trioxide, oleum, and concentrated sulfuric acid are considered together because of their close physical relationship and because they can, in certain cases, be used interchangeably. This group accounts for the preponderant production of aromatic sulfonates

Sulfur trioxide is theoretically the most efficient sulfonating and sulfating agent, because only direct addition is involved.

$$RH + SO_3 \rightarrow RSO_3H$$



$$ROH + SO_3 \rightarrow ROSO_3H$$

However, sulfur trioxide combines with water and aromatic compounds with the evolution of so much heat that its activity must be moderated For this reason, the hydrates of SO_3 (oleum and sulfuric acid) are generally used. Here, SO_3 is the true reactive species, the water functioning only as a complexing agent and a solvent An increase in the water content lowers the activity of the reagent for sulfonation; the reaction rate is inversely proportional to the square of the water concentration. When the SO_3 concentration in the sulfonating agent has been reduced to a critical level, which depends upon the organic compound being treated, sulfonation stops. The critical concentrations (π values) for the monosulfonation of naphthalene, benzene, and nitrobenzene are approximately 52, 64, and 82% SO_3.

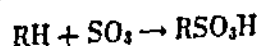
Catalysts. The addition of certain chemicals, usually in small amounts, can have a marked influence on some sulfonations. The addition of mercury changes the orientation in a number of aromatic sulfonations This is of great importance in the preparation of α-anthraquinone sulfonates. In the absence of mercury compounds, the β sulfonates are obtained exclusively. About 1% mercury, as metal or salt equivalent, based on anthraquinone used is required Mercury also affects the orientation in the sulfonation of benzoic acid, phthalic anhydride, and nitrobenzene. In these reactions the quantity of mercury needed is high, and the influence on orientation is only partial.

Aromatic sulfonation, like nitration and halogenation, is a typical electrophilic substitution reaction Sulfonation, however, differs from the other reactions in two respects; it is reversible, and in certain cases (for example, in the sulfonation of naphthalenes) temperature has an important influence on the position of the entering group.

Equipment. Cast iron is resistant to the action of sulfuric acid in the range 75–100% in strength over a fairly wide temperature range and has been standard material of construction for sulfonation kettles for many years especially for numerous dye intermediates and for aromatic hydrocarbons. However, it has poor tensile strength and is corroded by oleum or sulfur trioxide This fault may be controlled in oleum sulfonations by adding acid slowly to the material being sulfonated to keep the acid concentration below the corrosive level

The use of lined steel vessels combines low cost and high strength with good corrosion resistance. Commonly used linings include glass, enamel, lead and type 316 stainless steel.

Continuous operations involving special equipment are advantageous only where the reaction rate is fast and where the volume of production (as benzenesulfonic acid and dodecylbenzene sulfonate) is large and relatively steady. *See* ORGANIC SULFUR COMPOUND; SUBSTITUTION REACTION; U... PROCESSES. [P.H.G.]



occur in mineral springs), strontium sulfate (celestite, $SrSO_4$), lead sulfide (galena, PbS), zinc sulfide (zinc blende, ZnS), copper iron disulfide (chalcopyrite, $CuFeS_2$), iron disulfide (iron pyrites, FeS_2), and mercury sulfide (cinnabar, HgS). It also occurs in mineral springs as hydrogen sulfide (H_2S) and is found in plants and animals as a constituent of such substances as eggs, mustard, garlic, cabbage, horseradish, wool, and hair. It is also found in organic materials such as coal and petroleum, and has even been found in meteorites.

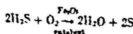
Preparation of the element. The extraction of sulfur is usually carried out by any of three methods. The most important is the Frasch process, developed in 1891 by Herman Frasch. Of lesser importance are the Sicilian method and a variation of the Claus method.

The Frasch process is used to extract sulfur from deposits such as those in Texas and Louisiana. It consists of boring a hole from the ground surface to the sulfur-bearing calcite deposit and lowering three pipes, concentrically arranged, to the ore bed (Fig. 1). Superheated water (165°C) is forced down the largest (6-in.) pipe into the ore bed where it melts the sulfur (melting point 112.8°C). Compressed hot air is pumped down the smallest (1-in.) pipe, and a frothy mixture of molten sulfur, water, and air is forced to the surface through the intermediate (3-in.) pipe. As it comes from the well, the sulfur has a purity of 99.5–99.9% and contains virtually no arsenic, selenium, or tellurium.

The Sicilian method consists of piling the sulfur-bearing rock into large mounds called *calcaroni*, which are ignited at the top. The heat of combustion of the sulfur in the ore causes underlying layers of sulfur to melt; this molten sulfur is poured into molds and is allowed to solidify. It often takes several months for a mound to be depleted, and only about 60% of the sulfur present in the ori...

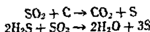
... usually refined by distillation. When the sulfur vapor is allowed to solidify directly on the walls of large masonry chambers, it is called flowers of sulfur because of the flowerlike designs in which it deposits. If the chambers are kept above the melting point of sulfur, the vapors condense to liquid sulfur, which is allowed to solidify in wooden molds. This form of the element is called roll sulfur.

Variations of the Claus method are sometimes used to obtain sulfur from gaseous hydrogen sulfide (H_2S), a by-product in the manufacture of many substances. It is based on the partial oxidation of the gas by oxygen in the air to give water and sulfur according to the equation.



$$2H_2S + O_2 \xrightarrow[\text{catalyst}]{Fe_2O_3} 2H_2O + 2S$$

Sulfur is also obtained as a by-product of many industrial processes by using coke or H_2S to reduce sulfur dioxide in flue gases:



$$SO_2 + C \rightarrow CO_2 + S$$
$$2H_2S + SO_2 \rightarrow 2H_2O + 3S$$

Properties of the element. The allotropes (different crystalline forms) of sulfur have been st...

modification of the element below 95.5°C (the transition point), and most of the other forms revert to this modification if allowed to stand below this temperature. The melting point of rhombic sulfur depends on the method of heating the substance and on the nature of the liquid sulfur with which it is in equilibrium. If rhombic sulfur is heated *very slowly*, it will convert to the monoclinic form...

ideally come into equilibrium with liquid sulfur only in the lambda form (λ-sulfur), and the melting point is 112.8°C. If the heating is rapid, the rhombic sulfur crystallizes from a melt in which λ-sulfur and mu-sulfur (μ-sulfur) are in equilibrium at 110°C. Rhombic sulfur is lemon-yellow, is

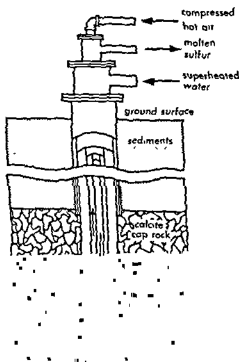


Fig. 1. Schematic diagram of the Frasch process for mining sulfur.

Hydroxy acids, such as $HOCH_2CH_2CH_2SO_3H$, form sultones analogous to the lactones of hydroxy-substituted carboxylic acids. Fatty acid esters of ethionic acid and amides of taurine find use as surface-active agents.

Aromatic sulfonic acids are made by sulfonation of aromatic compounds. Sulfuric acid, fuming sulfuric acid ($H_2SO_4 + SO_3$), chlorosulfonic acid ($ClSO_3H$), or sulfur trioxide may be used to introduce the sulfonic acid group. Aromatic sulfonic acids and their derivatives are important industrial chemicals. Of special value is their use as detergents, for example, sodium dodecylbenzenesulfonate. Sulfonated polymers act as cation-exchange resins and sulfonamide derivatives are valuable pharmaceuticals. The $—SO_3H$ group lends water solubility to many substances, hence increases their usefulness This application is particularly used in the manufacture of dyes and in some indicators, for example, Congo red or methyl orange.

Aromatic sulfonic acids also have applications as emulsifying agents, lubricating-oil additives, and rust inhibitors.

In synthesis, the SO_3H group can be replaced by Br or NO_2 groups by treating with bromine or nitric acid. The sodium salts, $ArSO_3^-Na^+$, yield phenols by fusing with alkali, and acidifying the melt. On fusion with sodium cyanide, they also yield nitriles,



$$ArSO_3Na + NaCN \rightarrow ArCN + Na_2SO_3$$

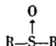
Sulfonation of aromatic hydrocarbons is a reversible process. Hence, treatment of $ArSO_3H$ with superheated steam removes the $—SO_3H$ group and is useful in purifying and separating aromatic hydrocarbons. The greatest utility of sulfonic acids as synthetic intermediates lies in their conversion to sulfonyl chlorides, which have a broad and useful reactivity See ORGANOSULFUR COMPOUND, SACCHARIN; SULFONAMIDE; SULFONATION; SULFONYL CHLORIDE; SULFURIC ACID [N.K.]

Sulfonyl chloride

One of a group of organosulfur compounds, RSO_2Cl, that are useful intermediates for synthesis and analysis They can be prepared by reactions of sulfonic acids, or their salts, with phosphorus chlorides, or by oxidative chlorination of thiols or disulfides The synthetic value lies in reductions to sulfinic acids, disulfides, or thiols, and in many displacement reactions of the chloride group· with amines. to sulfonamides, with thiols, to thiolsulfonate esters; with aromatic hydrocarbons, to sulfones (Friedel-Crafts reaction); with fluoride ion, to sulfonyl fluorides; and with alcohols, to sulfonate esters p-Toluenesulfonyl chloride (tosyl chloride) and p-nitrophenylazobenzenesulfonyl chloride are often used for making derivatives in qualitative organic analysis. The characteristic melting points of the products identify the unknown amine, alcohol, or thiol. See AMINE, ORGANOSULFUR COMPOUND [N.K.]

Sulfoxide

One of a group of organosulfur compounds

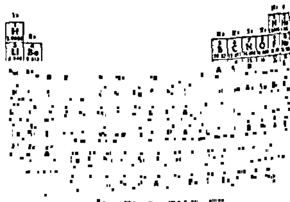


$$R—\overset{O}{\overset{\uparrow}{S}}—R$$

produced by controlled oxidation of organic sulfides (thioethers). The best-known example is dimethyl sulfoxide. If R and R′ are different, the sulfoxide can exist in two optically active forms Thus, the >SO bond is not a double bond (as in >C=O), because this would require a planar structure for sulfoxides. A pyramidal structure, with a dipolar $S^+ \rightarrow O^-$ bond, is assumed Sulfoxides are not as well known as sulfones, but an extensive chemistry of the sulfoxide group is rapidly building up. See OPTICAL ACTIVITY; ORGANOSULFUR COMPOUND; SULFONE (CHEMICAL). [N.K.]

Sulfur

Chemical element number 16, sulfur, S, has a chemical atomic weight of 32 066. The known stable isotopes and approximate per cent abundances in natural sulfur are S^{32}, (95.1%), S^{33},



(0.74%), S^{34}, (4.2%); S^{36}, (0 016%). The element was discovered prior to recorded history Its elemental character was first recognized by A. L Lavoisier in 1777

THE ELEMENT

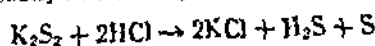
Natural occurrence. The abundance of sulfur in the earth's crust is 0 03–0 1%. It is often found as the free element near volcanic regions (impure deposits) in Japan, Sicily, and Mexico. Other deposits are located in New Zealand, Chile, Russia, Iceland, and Spain The largest known free sulfur deposits by far are in Texas and Louisiana and are associated with limestone and anhydrite caprock formations over salt domes. Other noteworthy deposits occur in California, Colorado, Wyoming Nevada, Utah, Mexico, and South America. Combined sulfur exists primarily in sulfates and sulfides such as calcium sulfate dihydrate (gypsum, $CaSO_4 \cdot 2H_2O$), barium sulfate (barite, $BaSO_4$), magnesium sulfate heptahydrate (epsom salt, $MgSO_4 \cdot 7H_2O$), sodium sulfate decahydrate (glauber's salt, $Na_2SO_4 \cdot 10H_2O$) (the last two usually

have been reported, but are far less well characterized than H_2S. *Hydrogen sulfide* can be prepared from the elements at elevated temperatures, although it is usually prepared by allowing a sulfide, such as iron sulfide, FeS, to be decomposed at room temperature by an acid such as hydrochloric acid. It is a colorless gas having a foul odor (similar to that of rotten eggs) and is considerably more poisonous than carbon monoxide, but warning of its presence (odor) is usually given before its concentration in the atmosphere is considered dangerous. Its density is 1.5392 g/liter at 0°C; melting point, −85.5°C, normal boiling point, −60.4°C. It is soluble in water, ethyl alcohol, carbon tetrachloride, and carbon disulfide. It burns in excess oxygen to form water and sulfur dioxide, in deficient oxygen to form water and free sulfur. The gas reacts directly with many metals to form the corresponding sulfides and with many nonmetals to form free sulfur. It is generally regarded as a good reducing agent. In water it behaves as a very weak acid (hydrosulfuric acid). It is used as a precipitating agent for many metal ions having insoluble or slightly soluble sulfides, and as a reducing agent (for example, in the preparation of hydrogen [illegible]

[illegible] rmal boiling point of 71°C. It is miscible with rbon disulfide, diethyl ether, and benzene, and is composed by water, acids, bases, and alcohol. ulfur dissolves in it to form hydrogen polysulfides ving properties similar to those of the disulfide.

Metal sulfides. Metal sulfides can be classified into three categories: acid sulfides (hydrosulfides, MHS where M = a univalent metal ion), normal sulfides (M_2S), and polysulfides (M_2S_x). The acid sulfides are soluble in water. Those normal sulfides which are soluble undergo hydrolysis in water to give the acid sulfide and usually hydrogen sulfide as well. The ease of hydrolysis of the soluble sulfides increases as the oxidation state of the metal ion increases. Most of the heavy metal sulfides are only very slightly soluble in water and are precipitated by hydrogen sulfide or ammonium sulfide. Metal acid sulfides and normal sulfides are usually prepared by reaction of the metal salt or hydroxide with hydrogen sulfide or ammonium sulfide, by reduction of the metal sulfates with hot carbon or by direct combination of the metal with hot sulfur. The alkali and alkaline earth sulfides are colorless, whereas the heavy metal sulfides are usually deeply colored. The soluble sulfides are used as reducing agents and in the preparation of sulfur-containing dyes, as depilatory materials and pesticides, and in the tanning of leather and the preparation of phosphors.

Polysulfides are formed by the reaction of free sulfur with solutions of alkali metal sulfides. The intensity of the [illegible] lfides generally increases with increasing uptake of sulfur. Metal disulfides and polysulfides (for example, K_2S_x; $x = 2–6$) undergo hydrolysis to a much smaller extent than normal sulfides and are decomposed by acids, usually with the deposition of free sulfur:



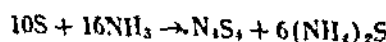
$$K_2S_2 + 2HCl \rightarrow 2KCl + H_2S + S$$

Polysulfides are used in analytical determinations of several metal ions. Organic derivatives of polysulfides such as dimethyl trisulfide, $(CH_3)_2S_3$, are known.

Other sulfides. Carbon-sulfur compounds and compounds containing the carbon-sulfur bond are well known. In addition to those previously described, some important compounds are: carbon disulfide, CS_2, a liquid which has a normal boiling point of 46.2°C and a melting point of −111.6°C, and which is an excellent solvent for elemental sulfur and phosphorus; carbon monosulfide, CS, an unstable gas formed by passing an electric discharge through carbon disulfide; and carbon oxysulfide, SCO, formed from carbon monoxide and free sulfur at an elevated temperature, having a normal boiling point of −50.2°C and a freezing point of −138.8°C.

Nitrogen-sulfur compounds which have been [illegible] greater electronegativity of nitrogen, although in the literature they are usually called sulfides.

Sulfur nitride is the best characterized of these. It is a yellow-to-red crystalline material which melts at about 178°C and sublimes at reduced pressures and elevated temperatures. It is soluble in carbon disulfide, benzene, ethanol, liquid ammonia, and carbon tetrachloride and reacts with water to form ammonium, sulfite, and pentathionate ions and free sulfur. It also reacts with chlorine to form $N_4S_4Cl_4$. It has the cradlelike structure of arsenic sulfide, As_4S_4 (realgar), with sulfur atoms in the arsenic positions and nitrogen atoms in the sulfur positions (Fig. 3). It can be prepared by the reaction of sulfur with liquid ammonia at or above −11.5°C:



$$10S + 16NH_3 \rightarrow N_4S_4 + 6(NH_4)_2S$$

The other nitrogen sulfides are of lesser importance. Phosph [illegible] certainly exceed [illegible] of P [illegible] that [illegible] in c [illegible] oxygen compounds (for example, alcohols) into the corresponding sulfur analogs. P_4S_{10} is used in the preparation of flotation agents for concentrating sulfide ores. P_4S_3 is used in the manufacture of matches. The compounds can be prepared from the elements. Phosphorus oxysulfide, $P_4S_4O_6$, is a colorless compound which melts at about 102°C and has a normal boil

insoluble in water, slightly soluble in ethyl alcohol, diethyl ether, and benzene, and very soluble in carbon disulfide. Its density is 2.06 g/cm^3, and its hardness is 2.5 on the Mohs scale. Its molecular formula is S_8, and its molecular configuration is a ring of eight covalently bonded sulfur atoms in the shape of a puckered crown, with bond distances of 2.12 A (A = angstrom unit = 10^{-8} cm) and bond angles of 105° (Fig. 2).

Monoclinic sulfur. Monoclinic sulfur, also called prismatic sulfur and beta-sulfur (β-sulfur), is the stable modification of the element above the transition temperature and below the melting point. It crystallizes from molten sulfur in needlelike prisms which are almost colorless It has a density of 1.96 g/cm^3 and a melting point of 119.3°C. Its molecular configuration is also an 8-atom puckered crown structure, and this form is also soluble in carbon disulfide and insoluble in water.

Plastic sulfur. Plastic sulfur, also called gamma-sulfur (γ-sulfur), is formed when molten sulfur at or near its normal boiling point is quenched to the solid state (such as by pouring it into cold water) This form of sulfur is amorphous and is only partially soluble in carbon disulfide. It is thought to be composed of two types of sulfur: λ-sulfur (S_λ, soluble in CS_2) and μ-sulfur (S_μ, insoluble in CS_2). These forms probably exist in liquid sulfur also, because plastic sulfur appears to be only the supercooled liquid After long standing at room temperature, this form of sulfur reverts to the rhombic form. Plastic sulfur exists as long zigzag chains of sulfur atoms, and if it is strongly stretched, it behaves like rubber because the zigzag chains are oriented in straight lines, and the material becomes fibrous and rigid

Purple sulfur. Purple sulfur is formed by the sudden cooling of sulfur vapor at an elevated temperature (where the sulfur exists as S_2) to −195°C. This purple modification is believed to be composed of S_2 units. It is unstable and reverts to yellow sulfur upon being warmed to room temperature.

Liquid sulfur. Liquid sulfur exhibits the remarkable property of increasing in viscosity as its temperature is raised. Its color becomes dark reddish-black as its viscosity increases and both color darkening and viscosity reach a maximum at 200°C. Above this temperature, the color lightens and the viscosity diminishes. It is thought that at the melting point, liquid sulfur is largely S_λ (S_8 rings, yellow), and that as the temperature increases, the percentage of S_μ (polymeric sulfur chains, reddish-black) also increases. At 200°C, there is thought to be a maximum of the highly polymeric sulfur chains which can intertwine to give maximum light absorption and viscosity. Above this temperature, the chains break, are reduced in length, and the viscosity decreases with increasing temperature There is also believed to be another form of sulfur in solid and molten sulfur called pi sulfur (S_π) but it has not been so extensively studied as the forms already mentioned, and little is known about it. A form of sulfur called rho-sulfur (S_ρ), believed to be S_6, has been reported as being obtained by the extraction of acidified aqueous sodium thiosulfate solution with toluene. Little is known of this form of sulfur, however.

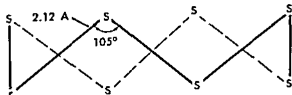


Fig 2. The structure of the S_8 molecule.

Gaseous sulfur. At the normal boiling point of the element (444.60°C) gaseous sulfur is orange-yellow in color. As the temperature is raised the color becomes deep red and then becomes lighter until at 650°C, it is straw-yellow. Several molecular species are in equilibrium in gaseous sulfur: S_8, S_6, S_4, and S_2, the proportions varying with the temperature. At the normal boiling point, the vapor is largely S_8; at 750°C, it is largely S_2; above 2000°C, it is largely dissociated into sulfur atoms

Other forms Milk of sulfur is a suspension of finely divided, amorphous sulfur in water, obtained by the decomposition of polysulfide solutions with acid. It is soluble in carbon disulfide Colloidal sulfur, also called delta-sulfur (δ-sulfur), is a colloidal dispersion of sulfur in water produced by the action of gaseous hydrogen sulfide on cold concentrated aqueous solutions of sulfur dioxide, or by decomposing sodium thiosulfate with dilute sulfuric acid. It dissolves quite slowly in carbon disulfide.

Chemical properties and uses. Sulfur is an active element which combines directly with most of the known elements It can exist in both positive and negative oxidation states, and can form ionic as well as covalent and coordinate covalent compounds.

The uses of sulfur are limited primarily to the manufacture of sulfur compounds. However, large quantities of elemental sulfur are used in the vulcanization of rubber, in lime-sulfur sprays to destroy plant parasites, in the manufacture of artificial fertilizer and certain types of cements and electric insulators, in certain ointments and medicinals, and in the manufacture of gunpowder and matches Sulfur compounds are used in the manufacture of chemicals, textiles, soaps, fertilizers, leather, plastics, refrigerants, bleaching agents, drugs, dyes, paints, paper, and many other products.

A test for the detection of elemental sulfur is the formation of a red solution when sulfur dissolves in piperidine All of the important allotropic forms show this behavior

PRINCIPAL COMPOUNDS

Sulfides. Hydrogen sulfide (H_2S) is the most [illegible]

...abilized forms of sulfur trioxide called Sulfans ... are called ... [illegible] ...oxide ap... between ... [illegible] ...onomeric and trimeric sulfur trioxide. It has a ...ormal boiling point of 44.5°C. Gaseous sulfur tri...xide is a monomer, and its structure is a planar ...quilateral triangle with the sulfur at the center. ... [illegible] ... 120° and the S—O ... [illegible] ... reactive

The γ form is the most reactive, and the α form the ...east. All forms react with water with the libera...tion of heat to produce sulfuric acid. The reaction with sulfuric acid, H_2SO_4, to form pyrosulfuric ... [illegible] ...

... or sulfonic acids upon reaction with organic materials. It reacts directly with basic metal oxides to form sulfates and with hydrogen chloride to form chlorosulfonic acid, HSO_3Cl. It is decomposed at high temperatures to sulfur dioxide and oxygen.

Sulfur trioxide is usually prepared by the catalytic oxidation of sulfur dioxide at 400–665°C. Vanadium pentoxide is the catalyst most often used although platinum metal, the sulfates of nickel and cobalt, and the oxides of iron, tungsten, molybdenum, and chromium can also be used. Small quantities of sulfur trioxide are prepared by the thermal decomposition of certain metal sulfates such as ferric sulfate or of pyrosulfates such as sodium pyrosulfate, $Na_2S_2O_7$. The trioxide is also prepared by the reaction between sulfur dioxide and ozone, O_3, at room temperature and by the reaction between nitric oxide and sulfur dioxide at high pressures and room temperature.

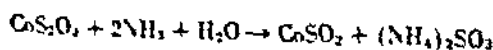


$$2NO + 2SO_2 \rightarrow 2SO_3 + N_2$$

Sulfur trioxide is used ... [illegible] ...

...ree acids themselves have not been isolated because of their instability. The table summarizes these materials.

Sulfoxylic acid. Best characterized as the cobalt(II), or cobaltous, salt, sulfoxylic acid precipitates as a brown solid upon treatment of sodium hyposulfite solution ($Na_2S_2O_4$) with cobalt(II) acetate and aqueous ammonia:



$$CoS_2O_4 + 2NH_3 + H_2O \rightarrow CoSO_2 + (NH_4)_2SO_3$$

The diethyl ester (C_2H_5—O—S—O—C_2H_5, boiling point 117°C at 733 mm Hg) is a colorless liquid which is not miscible with water. The sulfoxylates are extremely susceptible to oxidation.

Hyposulfurous acid. Hyposulfurous acid may be prepared in aqueous solution by the reduction of sulfurous acid solution with amalgamated zinc, but ... solution is unstable. The material behaves as a

Oxy acids of sulfur

Name and formula	Probable structure	Known forms
Sulfoxylic acid, H_2SO_2	HO—S—OH	Salts, esters
Hyposulfurous acid (dithionous, hydrosulfurous), $H_2S_2O_4$	HO—S(O)—S(O)—OH	Acid, salts
Sulfurous acid, H_2SO_3	HO—S(O)—OH	Acid, salts, esters
Thiosulfurous acid, $H_2S_2O_2$	HO—S(S)—OH	Esters
Pyrosulfurous acid, $H_2S_2O_5$	HO—S(O)—S(O)(O)—OH	Salts
Sulfuric acid, H_2SO_4	HO—S(O)(O)—OH	Acid, salts, esters
Pyrosulfuric acid, $H_2S_2O_7$	HO—S(O)(O)—O—S(O)(O)—OH	Acid, salts
Thiosulfuric acid, $H_2S_2O_3$	HO—S(O)(S)—OH	Salts
Dithionic acid, $H_2S_2O_6$	HO—S(O)(O)—S(O)(O)—OH	Acid, salts
Polythionic acids (tri-, tetra-, penta-, hexa-), $H_2S_xO_6$ (x = 3–6)	HO—S(O)(O)—$(S)_{x-2}$—S(O)(O)—OH (tentative)	Salts
Peroxymonosulfuric acid, H_2SO_5	H—O—O—S(O)(O)—OH	Acid, salts
Peroxydisulfuric acid, $H_2S_2O_8$	HO—S(O)(O)—O—O—S(O)(O)—OH	Acid, salts
Sulfenic acids, RSOH (R = alkyl or acyl group such as CH_3)	HO—S—R	Esters, halides
Sulfinic acids, RSO_2H	HO—S(O)—R	Acids, esters, halides
Sulfonic acids, RSO_3H	HO—S(O)(O)—R	Acids, esters, halides, amides
Thiosulfonic acids, RS_2O_2H	H—S—S(O)(O)—R	Salts, esters

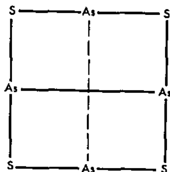


Fig 3. The structure of realgar, As_4S_4. This has a cradle structure in the vapor state, with two arsenic atoms above and two below the plane of the four sulfur atoms. Interatomic distances. As—S = 2.23 A, As—As = 2 49 A; bond angles: As—S—As = 101°, S—As—S = 93°.

ing point of 295°C. It is soluble in carbon disulfide and benzene and is prepared by a reaction between P_4S_{10} and P_4O_{10} at elevated temperatures.

Oxides. The oxides of sulfur which have been characterized have the formulas SO, S_2O_3, SO_2, SO_3, S_2O_7, and SO_4. Sulfur dioxide (SO_2) and sulfur trioxide (SO_3) are of far greater importance than the others. Sulfur monoxide (SO) can be prepared by passing an electric discharge through a mixture of sulfur vapor and sulfur dioxide at low temperatures. It is a gas at ordinary temperatures and produces an orange-red deposit when cooled to the temperature of liquid air. The gas is stable only at reduced pressures, is probably dimeric, and is soluble in thionyl chloride. The solid is probably a long-chain polymer. Sulfur sesquioxide (S_2O_3) is a blue-green solid which is stable only below 15°C It is prepared by a reaction of free sulfur with excess liquid sulfur trioxide, and it reacts with water to produce free sulfur, sulfurous, sulfuric, and several thionic acids It appears to be a high polymer, although its structure is not known as yet. Sulfur heptoxide (S_2O_7) is a poorly characterized material which can be prepared by passing sulfur dioxide or sulfur trioxide and oxygen or ozone through an electric discharge. Its structure is not known, although it is believed that a peroxide group (—O—O—) must exist in the compound Sulfur tetroxide (SO_4) is prepared by passing a mixture of sulfur dioxide and excess oxygen at reduced pressure through a glow discharge The product is a white solid melting at 3°C (with decomposition) and is a monomer. It is a strong oxidizing agent, and its structure has not yet been [illegible]

[illegible]

less gas with a pungent odor, melting at −75 46°C. and boiling at −10.02°C It has an angular structure, the bond angle (O—S—O) being 119° and the sulfur-oxygen bond distances being 1.43 A. Sulfur dioxide can act as an oxidizing agent (for example, toward hydrogen sulfide, hydrogen, and carbon monoxide) and as a reducing agent (for example, toward permanganate ion). It reacts with water giving an acidic solution (often called sulfurous acid) and bisulfite (HSO_3^-) and sulfite (SO_3^{--}) ions. The equilibrium constant (18°C) for the dissociation to the acid sulfite (HSO_3^-) is 1.6×10^{-2} and for the dissociation of the acid sulfite to sulfite it is 1.0×10^{-7} at the same temperature. At low temperatures, sulfur dioxide forms solvates with metal fluorides, iodides, and thiocyanates ($NaI{\cdot}4SO_2$). The dioxide is used as a refrigerant gas because of the ease with which it is liquefied and because of its relatively high heat of vaporization. It is also used as a disinfectant and preservative because of its germicidal properties. It also finds use as a bleaching agent and in the refining of petroleum products Its major use, however, is in the manufacture of sulfur trioxide and sulfuric acid.

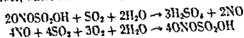
Be [illegible]
also [illegible]
prop[illegible]
and [illegible]
thionyl sulfite, (SO)SO_3. Liquid sulfur dioxide is partially miscible with water and also forms crystalline hydrates with it (for example, $SO_2{\cdot}H_2O$); it is completely miscible with benzene Because liquid sulfur dioxide is not a reducing agent, oxidation reactions such as halogenations can be carried out in it. The sulfites of metal ions whose hydroxides are amphoteric in water (for example Al^{III}) are amphoteric in liquid sulfur dioxide.

Sulfur dioxide is usually prepared by burning sulfur or roasting metallic sulfides in air or oxygen by the reaction of metals such as copper with concentrated sulfuric acid at elevated temperatures, or by the reaction between a sulfite (Na_2SO_3) or acid sulfite ($NaHSO_3$) and a strong acid (H_2SO_4)

Sulfur trioxide Sulfur trioxide (SO_3) exists in several forms, and their relationships are still not completely understood. Gamma-sulfur trioxide (γ-SO_3) is trimeric and resembles ice in appearance. Its equilibrium melting point is 16 8°C and it can be prepared by condensing extremely dry sulfur trioxide vapor at −80°C Beta-sulfur trioxide (β-SO_3) is an asbestoslike polymeric substance in which the SO_4 groups are linked in long chains. Its equilibrium melting point is 32 5°C It is prepared in a manner similar to the α form using ordinary sulfur trioxide (which contains some moisture) The condensation yields a mixture of the β and γ forms, and the γ form is removed by distillation Alpha-sulfur trioxide (α-SO_3) is also an asbestoslike solid which is similar to the β form except that the SO_4 chains are also joined in a layer type of structure Its equilibrium melting point is 62 3°C and it can be formed by condensing gaseous sulfur trioxide at liquid air temperatures The γ and β forms are metastable with respect to the α form, and conversion to the α form is catalyzed by traces of moisture However, this conversion to a polymer is inhibited by sulfur, tellurium carbon tetrachloride, and phosphorus oxytrichloride, which are used i[illegible]ercially as [illegible]

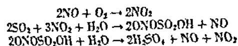
chröder Grillo, Mannheim, and Tenteleff processes

In the lead-chamber process, nitrogen oxides are introduced into the sulfur dioxide-air mixture which is passed upward through a tower about 25 ft high (called the Glover tower), in which it is sprayed with a sulfuric acid-nitrosyl sulfuric acid ($ONOSO_2OH$) mixture from the Gay-Lussac tower, where the following reactions are probable:



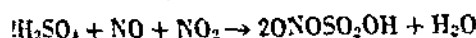
$$2ONOSO_2OH + SO_2 + 2H_2O \rightarrow 3H_2SO_4 + 2NO$$
$$4NO + 4SO_2 + 3O_2 + 2H_2O \rightarrow 4ONOSO_2OH$$

The gases (a mixture of SO_2, O_2, NO, and NO_2) then pass through several lead chambers and are sprayed with steam. The reactions which occur here are not fully understood, but it is believed that they take place as follows



$$2NO + O_2 \rightarrow 2NO_2$$
$$2SO_2 + 3NO_2 + H_2O \rightarrow 2ONOSO_2OH + NO$$
$$2ONOSO_2OH + H_2O \rightarrow 2H_2SO_4 + NO + NO_2$$

The gases then enter the Gay-Lussac tower where they are sprayed with fairly concentrated sulfuric acid (about 80%), which results in the formation of additional nitrosyl sulfuric acid



$$2H_2SO_4 + NO + NO_2 \rightarrow 2ONOSO_2OH + H_2O$$

...ich is then pumped to the Glover tower to complete the cycle. The nitrogen oxides act as catalysts and because there is some loss, they are replaced periodically by the catalytic oxidation of ammonia. The acid is tapped from the Glover tower and the lead chambers, and is quite impure, containing oxides of arsenic and nitrogen, and many metal salts. Lead chamber acid is used primarily in the manufacture of fertilizer because removal of these impurities is not essential (Fig 5).

The chemical properties of sulfuric acid are of considerable importance, and because of them, sulfuric acid has become the largest tonnage manufactured chemical in the world. In 1957, production in the United States alone was more than 16,000,000 short tons.

The compound is a strong acid in water, and reacts with most metals in either the dilute or concentrated form. Iron and steel do not react with the concentrated acid, which allows it to be shipped in tank cars. The concentrated acid is a strong oxidizing agent, especially at elevated temperatures, and will react with metals, carbon, sulfur and other oxidizable materials. Because of its relatively high boiling point (338°C for the 98.3% acid), it can react with salts at elevated temperatures to liberate volatile acids such as HCl. The concentrated acid is a strong dehydrating agent, and reacts vigorously with water with the evolution of much heat (205 cal/g of H_2SO_4). It will also extract hydrogen and oxygen (to form water) from organic materials such as sugar, wood, and animal tissues, thereby decomposing them and leaving carbon.

The uses of sulfuric acid are varied. It finds primary use in the manufacture of superphosphate

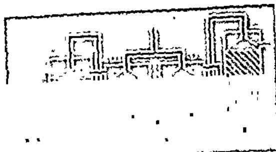


Fig 5 The lead chamber process A, sulfur or pyrite burners, B, inlet for nitrogen oxides, C, Glover tower; D, lead chambers, E, Gay-Lussac tower, P, pumps

fertilizers. Petroleum refining also consumes large quantities of the acid, as does the manufacture of many chemicals, including sulfates, hydrochloric and nitric acids, dyes, drugs, and explosives. The iron and steel, storage battery, paint, plastic, metallurgical, and textile industries also use large quantities of the acid. It is sometimes used as a solvent in chemical research because of its strong hydrogen-ion donating ability. Its use is so widespread that its tonnage production is often used as an indicator of general business conditions in the country.

Both normal sulfates such as Na_2SO_4 and acid sulfates (bisulfates or hydrogen sulfates) such as $NaHSO_4$ are well known. The structure of the sulfate ion is tetrahedral, with the sulfur at the center and an oxygen at each corner. Most sulfates of substances can be prepared by treating the substance with dil...

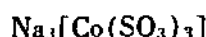
...ue exceptions are certain alkaline-earth and lead sulfates which are sparingly soluble). Most normal sulfates are thermally stable except at extremely high temperatures. Acid sulfates are converted to pyrosulfates and normal sulfates by intense heat. The sulfate ion can act as a ligand (coordinating agent) in coordination compounds such as $[Co(NH_3)_5SO_4]Cl$. Organic sulfates are known, for example, diethyl sulfate, $(C_2H_5)_2SO_4$, mp, −24.5°C, normal bp, 208°C, which is used as an alkylating agent. Double sulfates form quite easily the two ...
ing th...
(M^I = ... univalent positive ion) and the schönites, $M^I_2M^{II}(SO_4)_2 \cdot 6H_2O$. Sulfates are usually determined analytically by treatment with aqueous barium chloride solution (acidified with hydrochloric acid) to give a precipitate of barium sulfate which is washed, filtered, dried, and weighed.

Pyrosulfuric acid. Pyrosulfuric acid is the product of the reaction of equimolar quantities of pure sulfuric acid and sulfur trioxide. Its melting point is 35.15°C. It is an excellent sulfonating agent and loses sulfur trioxide on being heated. It also reacts vigorously with water, liberating a considerable quantity of heat. Alkali metal pyrosulfates

strong acid (for the first hydrogen), and the solution is a good reducing medium. Hyposulfites (dithionites or hydrosulfites) are more stable as dry solids than as the acid solution. They are also strong reducing agents, and can be prepared by reduction of the corresponding metal bisulfite solution with zinc dust, by electrolytic reduction of the sulfite, or by treatment of an active metal amalgam such as sodium amalgam with dry sulfur dioxide. Metal hyposulfites are used primarily as reducing agents in the dye industry.

Sulfurous acid. Sulfurous acid is not actually known as a pure substance, although the hydrate $SO_2 \cdot 7H_2O$ can be crystallized from concentrated aqueous solutions of sulfur dioxide at low temperatures. Aqueous solutions of sulfur dioxide contain primarily hydrated protons, bisulfite ions (HSO_3^-), and a much smaller concentration of sulfite ions (SO_3^{--}). They are strongly reducing, and can be oxidized to sulfate and dithionate. Such solutions can also behave as oxidizing agents in the presence of strong reductants, such as iodide ion and zinc. Both normal sulfites (Na_2SO_3) and acid (hydrogen) sulfites ($NaHSO_3$) are well known.

The structure of the sulfite ion is pyramidal, with one atom at each corner. Of the normal sulfites, only the alkali metal salts are appreciably soluble, although many metal bisulfites (acid sulfites) are soluble. Both normal and acid sulfites are usually prepared by treatment of the corresponding carbonate or hydroxide with appropriate quantities of sulfur dioxide, and they both liberate sulfur dioxide on treatment with excess acid. Solutions of sulfites dissolve free sulfur to form thiosulfates. Bisulfites can form addition compounds with many organic compounds. The sulfite ion forms coordination compounds with metal ions, for example,



$$Na_3[Co(SO_3)_3]$$

as well as esters, such as $(CH_3O)_2SO$ (normal boiling point 121.5°C), which are good alkylating agents. Sulfites are used extensively as reducing agents, as addition agents to organic compounds, and in the manufacture of paper from wood.

Thiosulfurous acid. Thiosulfurous acid is known only in the form of its salts, and even these are not very well characterized. They can be formed by the action of a dry sodium alkylate such as $NaOCH_3$ on sulfur monochloride (S_2Cl_2) in ligroin solution. The boiling point of the dimethyl compound is 33°C at 15 mm Hg, and of the diethyl compound, 67°C at 16 mm Hg. A typical structure postulated for these esters is (for the dimethyl ester) $H_3C—O—S—S—O—CH_3$. These substances are stable toward oxidation by atmospheric oxygen and are hydrolyzed by strong bases to give thiosulfates and sulfur.

Pyrosulfurous acid. Pyrosulfurous acid is known only in the form of pyrosulfites, which are usually prepared from aqueous solutions of alkali metal sulfites and sulfur dioxide or by heating alkali metal acid sulfites. The sodium salt is used primarily in the dye, printing, and photographic industries.

Sulfuric acid. This is a colorless, viscous liquid whose melting point is 10.31°C. When this liquid is heated, it gives off sulfur trioxide and begins to boil at 290°C. However, the normal boiling point increases until it reaches 317°C, at which point the acid is 98.54% H_2SO_4. Gaseous H_2SO_4 begins to dissociate into sulfur trioxide and water vapor at about 300°C, the dissociation being 50% complete at 350°C, and essentially complete at 444°C. The acid forms the hydrates $H_2SO_4 \cdot H_2O$ (melting point, 8.47°C), $H_2SO_4 \cdot 2H_2O$ (mp, −39.46°C), and $H_2SO_4 \cdot 4H_2O$ (mp, −28.25°C). The structure of the H_2SO_4 molecule is a tetrahedron with a sulfur atom at the center and two OH groups and two oxygen atoms at the corners. The sulfur-oxygen bond distances are 1.51 A. The density of 100% H_2SO_4 is 1.8384 g/ml at 15°C. The acid dissociates essentially completely in water to give a hydrated proton and the acid sulfate (bisulfate) ion. The bisulfate ion dissociates to a large degree in dilute aqueous solution ($pK_2 = 1.7$ at 25°C) to give the normal sulfate anion and another hydrated proton.

The preparation of sulfuric acid is usually carried out by the contact process. A far less important process today is the lead chamber process because this procedure gives relatively dilute acid (60–78%) which has limited usefulness, whereas the contact process can give acid of any desired concentration.

In the contact process, sulfur is burned or iron pyrites roasted in air to produce sulfur dioxide, which is then oxidized to sulfur trioxide in the presence of a suitable catalyst (usually vanadium pentoxide or platinum). The sulfur trioxide is absorbed in concentrated sulfuric acid to produce oleum (pyrosulfuric acid, $H_2S_2O_7$), which is then treated with water to produce sulfuric acid of any desired concentration. The schematic diagram of the process shows that the burner gas (about 25% sulfur dioxide, 30% oxygen, and 45% nitrogen) is filtered to remove dust particles, and is cleaned, dried, and treated to remove arsenic (a catalyst poison) before being heated and converted to sulfur trioxide on the catalyst bed (Fig. 4). There are many variations, for example, the Badische

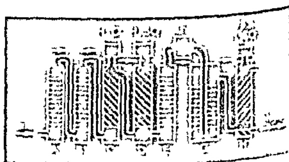
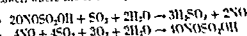


Fig. 4 The contact process. A, sulfur burner; B, dust removal tower; C and D, scrubbing towers to clean sulfur dioxide; E, arsenic removal tower; F, heater; G, catalyst chamber; H, sulfur trioxide absorber tower.

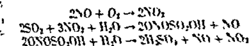
chröder Grillo, Mannheim, and Tenteleff processes

In the lead-chamber process, nitrogen oxides are ntroduced into the sulfur dioxide-air mixture which is passed upward through a tower about 25 t high (called the Glover tower), in which it is prayed with a sulfuric acid nitrosyl sulfuric acid (ONOSO$_2$OH) mixture from the Gay-Lussac ower, where the following reactions are probable:



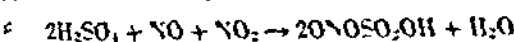
$$2ONOSO_2OH + SO_2 + 2H_2O \rightarrow 3H_2SO_4 + 2NO$$
$$4NO + 4SO_2 + 3O_2 + 2H_2O \rightarrow 4ONOSO_2OH$$

The gases (a mixture of SO_2, O_2, NO, and NO_2) then pass through several lead chambers and are sprayed with steam. The reactions which occur there are not fully understood, but it is believed that they take place as follows:



$$2NO + O_2 \rightarrow 2NO_2$$
$$2SO_2 + 3NO_2 + H_2O \rightarrow 2ONOSO_2OH + NO$$
$$2ONOSO_2OH + H_2O \rightarrow 2H_2SO_4 + NO + NO_2$$

The gases then enter the Gay-Lussac tower where they are sprayed with fairly concentrated sulfuric acid (about 80%), which results in the formation of additional nitrosyl sulfuric acid.



$$2H_2SO_4 + NO + NO_2 \rightarrow 2ONOSO_2OH + H_2O$$

hich is then pumped to the Glover tower to complete the cycle. The nitrogen oxides act as catalysts, and because there is some loss, they are replaced periodically by the catalytic oxidation of mmonia. The acid is tapped from the Glover ower and the lead chambers, and is quite impure, containing oxides of arsenic and nitrogen, and many metal salts. Lead chamber acid is used primarily in the manufacture of fertilizer because removal of these impurities is not essential (Fig 5)

The chemical properties of sulfuric acid are of considerable importance and because of them, sulfuric acid has become the largest tonnage manufactured chemical in the world In 1957, production in the United States alone was more than 16,000,000 short tons

The compound is a strong acid in water, and reacts with most metals in either the dilute or concentrated form Iron and steel do not react with the concentrated acid, which allows it to be shipped in tank cars The concentrated acid is a strong oxidizing agent, especially at elevated temperatures, and will react with metals carbon, sulfur and other oxidizable materials. Because of its relatively high boiling point (338°C for the 98.3% acid), it can react with salts at elevated temperatures to liberate volatile acids such as HCl The concentrated acid is a strong dehydrating agent, and reacts vigorously with water with the evolution of much heat (205 cal/g of H_2SO_4) It will also extract hydrogen and oxygen (to form water) from organic materials such as sugar, wood, and animal tissues, thereby decomposing them and leaving carbon.

The uses of sulfuric acid are varied It finds primary use in the manufacture of superphosphate

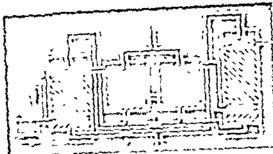


Fig 5 The lead chamber process. A, sulfur or pyrite burners, B, inlet for nitrogen oxides, C, Glover tower; D, lead chambers, E, Gay-Lussac tower; F, pumps.

fertilizers. Petroleum refining also consumes large quantities of the acid, as does the manufacture of many chemicals, including sulfates, hydrochloric and nitric acids, dyes, drugs, and explosives. The iron and steel, storage battery, paint, plastic, metallurgical, and textile industries also use large quantities of the acid It is sometimes used as a solvent in chemical research because of its strong hydrogen ion donating ability Its use is so widespread that its tonnage production is often used as an indicator of general business conditions in the country

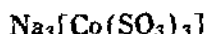
Both normal sulfates such as Na_2SO_4 and acid sulfates (bisulfates or hydrogen sulfates) such as $NaHSO_4$ are well known The structure of the sulfate ion is tetrahedral, with the sulfur at the center and an oxygen at each corner Most sulfates of substances can be prepared by treating the substance with dilute or concentrated sulfuric acid or by oxidizing their sulfites or sulfides. Most normal and acid sulfates are quite soluble in water (notable exceptions are certain alkaline-earth and lead sulfates which are sparingly soluble) Most normal sulfates are thermally stable except at extremely high temperatures Acid sulfates are converted to pyrosulfates and normal sulfates by intense heat. The sulfate ion can act as a ligand (coordinating agent) in coordination compounds such as $[Co(NH_3)_5SO_4]Cl$. Organic sulfates are known, for example, diethyl sulfate, $(C_2H_5)_2SO_4$, mp, −24.5°C, normal bp, 208°C, which is used as an alkylating agent. Double sulfates form quite easily, the two most important series being the alums having the general formula $M^{I}M^{III}(SO_4)_2 \cdot 12H_2O$ (M^{I} = monovalent positive ion, M^{III} = trivalent positive ion) and the schönites, $M^{I}_2M^{II}(SO_4)_2 \cdot 6H_2O$ Sulfates are usually determined analytically by treatment with aqueous barium chloride solution (acidified with hydrochloric acid) to give a precipitate of barium sulfate which is washed, filtered, dried, and weighed

Pyrosulfuric acid. Pyrosulfuric acid is the product of the reaction of equimolar quantities of pure sulfuric acid and sulfur trioxide. Its melting point is 35 15°C. It is an excellent sulfonating agent and loses sulfur trioxide on being heated. It also reacts vigorously with water, liberating a considerable quantity of heat. Alkali metal pyrosulfates

strong acid (for the first hydrogen), and the solution is a good reducing medium Hyposulfites (dithionites or hydrosulfites) are more stable as dry solids than as the acid solution They are also strong reducing agents, and can be prepared by reduction of the corresponding metal bisulfite solution with zinc dust, by electrolytic reduction of the sulfite, or by treatment of an active metal amalgam such as sodium amalgam with dry sulfur dioxide. Metal hyposulfites are used primarily as reducing agents in the dye industry.

Sulfurous acid. Sulfurous acid is not actually known as a pure substance, although the hydrate $SO_2 \cdot 7H_2O$ can be crystallized from concentrated aqueous solutions of sulfur dioxide at low temperatures. Aqueous solutions of sulfur dioxide contain primarily hydrated protons, bisulfite ions (HSO_3^-), and a much smaller concentration of sulfite ions (SO_3^{--}). They are strongly reducing, and can be oxidized to sulfate and dithionate. Such solutions can also behave as oxidizing agents in the presence of strong reductants, such as iodide ion and zinc. Both normal sulfites (Na_2SO_3) and acid (hydrogen) sulfites ($NaHSO_3$) are well known.

The structure of the sulfite ion is pyramidal, with one atom at each corner. Of the normal sulfites, only the alkali metal salts are appreciably soluble, although many metal bisulfites (acid sulfites) are soluble Both normal and acid sulfites are usually prepared by treatment of the corresponding carbonate or hydroxide with appropriate quantities of sulfur dioxide, and they both liberate sulfur dioxide on treatment with excess acid. Solutions of sulfites dissolve free sulfur to form thiosulfates. Bisulfites can form addition compounds with many organic compounds. The sulfite ion forms coordination compounds with metal ions, for example,



$$Na_3[Co(SO_3)_3]$$

as well as esters, such as $(CH_3O)_2SO$ (normal boiling point 121 5°C), which are good alkylating agents. Sulfites are used extensively as reducing agents, as addition agents to organic compounds, ... of paper from wood

... known

... are not very well characterized. They can be formed by the action of a dry sodium alkylate such as $NaOCH_3$ on sulfur monochloride (S_2Cl_2) in ligroin solution The boiling point of the dimethyl compound is 33°C at 15 mm Hg, and of the diethyl compound, 67°C at 16 mm Hg. A typical structure postulated for these esters is (for the dimethyl ester) $H_3C—O—S—S—O—CH_3$. These substances are stable toward oxidation by atmospheric oxygen and are hydrolyzed by strong bases to give thiosulfates and sulfur.

Pyrosulfurous acid. Pyrosulfurous acid is known only in the form of pyrosulfites, which are usually prepared from aqueous solutions of alkali metal sulfites and sulfur dioxide or by heating alkali metal acid sulfites. The sodium salt is used primarily in the dye, printing, and photographic industries.

Sulfuric acid. This is a colorless, viscous liquid whose melting point is 10 31°C. When this liquid is heated, it gives off sulfur trioxide and begins to boil at 290°C. However, the normal boiling point increases until it reaches 317°C, at which point the acid is 98.54% H_2SO_4. Gaseous H_2SO_4 begins to dissociate into sulfur trioxide and water vapor at about 300°C, the dissociation being 50% complete at 350° ... complete at 411°C

The acid fo ...

point, 8.47° ...

$H_2SO_4 \cdot 4H_2O$ (mp, −28.25°C). The structure of the H_2SO_4 molecule is a tetrahedron with a sulfur atom at the center and two OH groups and two oxygen atoms at the corners. The sulfur-oxygen bond distances are 1.51 A. The density of 100% H_2SO_4 is 1.8384 g/*ml* at 15°C. The acid dissociates essentially completely in water to give a hydrated proton and the acid sulfate (bisulfate) ion The bisulfate ion dissociates to a large degree in dilute aqueous solution (pK_2 = 1.7 at 25°C) to give the normal sulfate anion and another hydrated proton.

The preparation of sulfuric acid is usually carried out by the contact process. A far less important process today is the lead-chamber process because this procedure gives relatively dilute acid (60–78%) which has limited usefulness, whereas the contact process can give acid of any desired concentration.

In the contact process, sulfur is burned or iron pyrites roasted in air to produce sulfur dioxide, which is then oxidized to sulfur trioxide in the presence of a suitable catalyst (usually vanadium pentoxide or platinum). The sulfur trioxide is absorbed in concentrated sulfuric acid to produce oleum (pyrosulfuric acid, $H_2S_2O_7$), which is then treated with water to produce sulfuric acid of any desired concentration The schematic diagram of the process shows that the burner gas (about 25% sulfur dioxide, 30% oxygen, and 45% nitrogen) is filtered to remove dust particles, and is cleaned, dried, and treated to remove arsenic (a catalyst poison) before being heated and converted to sulfur trioxide on the catalyst bed (Fig. 4). There are many variations, for example, the Badische

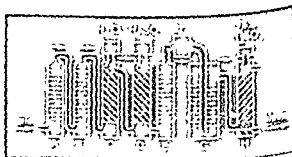


Fig. 4. The contact process A, sulfur burner; B, dust removal tower; C and D, scrubbing towers to clean sulfur dioxide; E, arsenic removal tower; F, heater; G catalyst chamber; H, sulfur trioxide absorber tower

Miscellaneous compounds. Other important organic oxygen-sulfur-containing compounds include the sulfoxides, R_2SO (which may be considered as being derived from sulfurous acid), and the sulfones, R_2SO_2 (from sulfuric acid). Aliphatic sulfoxides are usually prepared by the oxidation of sulfides with nitric acid or hydrogen peroxide, and the aromatic derivatives from aromatic hydrocarbons and sulfur dioxide or thionyl chloride in the presence of aluminum chloride. They are usually low-melting [illegible]

fuming nitric acid or permanganates; aromatic sulfones, by the action of sulfur trioxide on aromatic hydrocarbons or by the reaction of sulfonic acids with benzene and phosphorus(V) oxide at elevated temperatures. They are stable colorless solids which can be distilled without decomposition, for example, diphenyl sulfone, $(C_6H_5)_2SO_2$, mp 76°C, normal bp 379°C. Certain disulfones formed by the condensation of ketones and mercaptans followed by oxidation have medicinal value, for example, $(CH_3)_2C(SO_2C_2H_5)_2$, the hypnotic agent sulfonal.

The oxy halides of sulfur may be classified as derivatives of sulfoxylic acid, sulfurous acid (thionyl derivatives), and sulfuric acid (sulfuryl derivatives). Aryl sulfur halides (considered to be sulfoxylic acid derivatives) can be prepared from aryl mercaptans and halogens at low temperatures. An example is phenyl sulfur chloride (C_6H_5SCl), a red oil which boils at 149°C (12 mm Hg). Thionyl halides, such as SOF_2, $SOCl_2$, $SOBr_2$, and $SOClF$, are all known to have a triangular pyramidal structure with atoms at the corners [illegible] point of −40°C and a normal boiling point of 69.3°C. They are much more stable and less reactive than the corresponding thionyl derivatives, however. A pyrosulfuryl halide ($S_2O_5Cl_2$, pyrosulfuryl chloride) has also been characterized. It melts at −37.5°C and boils at 152.5°C (766 mm Hg).

Other important halogen derivatives of sulfuric acid are the organic sulfonyl halides and the halosulfonic acids. Alkyl and aryl sulfonyl halides, RSO_2X (X = F, Cl, Br), are colorless liquids or solids which usually have high boiling points (phenyl [illegible] acid), has a normal boiling point of 162.6°C and can be prepared from KHF_2 and oleum at elevated temperatures. Chlorosulfonic acid $HOSO_2Cl$ (mp −80°C, normal bp 151°C), can be prepared from hydrogen chloride and sulfur trioxide or oleum. It is decomposed extremely violently by water.

Halogen sulfur compounds which have been well characterized are S_2F_2 (sulfur monofluoride), SF_2, SF_4, SF_6, S_2F_{10}, S_2Cl_2 (sulfur monochloride), SCl_2, SCl_4, and S_2Br_2 (sulfur monobromide). These compounds have low melting points and low boiling points (S_2F_2, mp −120.5°C, normal bp −38.4°C) which hydrolyze in water (except for SF_6 and S_2F_{10}). Sulfur tetrafluoride is a remarkably selective fluorinating agent for organic compounds. Sulfur hexafluoride is quite inert, and this gas is used as a high voltage insulator. It melts at −50.8°C, sublimes at −63.7°C, and its structure is an octahedron with the sulfur at the center and a fluorine at each corner. The sulfur chlorides are used in the commercial manufacture of rubber and the monochloride, which is a liquid at room temperature, is also used as a solvent for organic compounds, sulfur, iodine, and certain metal compounds. These halides are usually prepared by direct combination of the elements. See ORGANOSULFUR COMPOUND; SELENIUM; TELLURIUM. [N.K.]

Bibliography: J. W. Mellor, *A Comprehensive Treatise on Inorganic and Theoretical Chemistry*, vol. 10, 1930; H. Remy, *Treatise on Inorganic Chemistry*, vol. 1, 1956; N. V. Sidgwick, *The Chemical Elements and Their Compounds*, vol. 2, 1950.

Sulfuric acid

A strong mineral acid with the chemical formula H_2SO_4. It is a colorless, oily liquid, sometimes called oil of vitriol or vitriolic acid. The pure acid has a density of 1.834 at 25°C, and freezes at 10.5°C. It is an important industrial commodity, used extensively in petroleum refining and in the manufacture of fertilizers, paints, pigments, dyes, and explosives.

Sulfuric acid is produced on a large scale by two commercial processes, the contact process and the lead-chamber process. In the contact process, sulfur dioxide (SO_2) is produced by burning sulfur or a sulfide such as that of iron, FeS_2, in air. The sulfur dioxide is converted to sulfur trioxide (SO_3) by reaction with oxygen in the presence of a catalyst such as platinized asbestos. Sulfuric acid is produced by the reaction of the sulfur trioxide with water. The lead chamber process depends upon the oxidation of sulfur dioxide by nitric acid in the presence of water, the reaction being carried out in large lead rooms.

The commercial acid may be concentrated to 98.3% by distillation, and pure sulfuric acid obtained by fractional crystallization. Sulfuric acid reacts vigorously with water to form several hydrates, of which the monohydrate, $H_2SO_4 \cdot H_2O$, is relatively stable. The concentrated acid, therefore, acts as an efficient drying agent, taking up moisture from the air and even abstracting the elements of water from such compounds as sugar and starch. Because of the formation of hydrates, the mixing of sulfuric acid and water is accompanied by the evolution of a great amount of heat.

such as $Na_2S_2O_7$ can be prepared by heating alkali metal acid sulfates or by a reaction between normal alkali metal sulfates and sulfur trioxide.

Thiosulfuric acid. Thiosulfuric acid is known only in its normal salts. The salts are stable only in the solid state or in neutral or alkaline solution. Thiosulfates are usually prepared by allowing free sulfur to dissolve in a solution of a metal sulfite, by the controlled oxidation of sulfides, or by the action of alkalies on polythionates. Thiosulfates are unstable in acid solution and decompose to free sulfur, pentathionates, and sulfites. The structure of the thiosulfate anion is analogous to that of the sulfate ion, one oxygen atom of the latter being replaced by a sulfur atom. Hydrated sodium thiosulfate (hypo) is used in the photographic industry as a fixing agent to dissolve unchanged silver salts from films and plates. It is also used as an antichlor to remove chlorine from bleached fabrics Thiosulfate ion acts as a coordinating agent in certain metal coordination compounds such as $Na_3[Ag(S_2O_3)_2]$, and many heavy metal thiosulfates which are ordinarily insoluble will dissolve in a solution containing excess thiosulfate ion because of the formation of a soluble complex Thiosulfate is determined analytically by titration with standard iodine solutions (or standard permanganate solutions which are used to liberate iodine from iodides). The reaction, which gives tetrathionate ion, is:



$$2S_2O_3^{--} + I_2 \rightarrow S_4O_6^{--} + 2I^-$$

and the titration is usually carried out until the blue color produced by starch in the presence of free iodine is just destroyed.

Thionic acids. The thionic acids are known as the salts only, except for dithionic acid, of which both the free acid and salts are known. The only structures known with certainty are those for the dithionate anion $(O_3S—SO_3)^{--}$ and trithionate anion $(O_3S—S—SO_3)^{--}$ (the sulfur atoms are nonlinear), but it is assumed that in the other polythionates, additional sulfur atoms are bonded to the central sulfur of the trithionate structure. Dithionic acid and dithionates may be prepared from sulfurous acid and sulfite solutions by oxidation with such oxidants as manganese dioxide, permanganates, and ferric or cobaltic hydroxides Dithionic acid is stable in dilute solution at room temperature, but undergoes decomposition on being heated The salts are stable in solution, and all of them appear to be soluble. Polythionates are definitely [illegible] hexathionate, and [illegible] They are [illegible] ıg agents, being oxidized to sulfate. They can be prepared by treatment of an aqueous solution of a thiosulfate with sulfur dioxide in the presence of arsenic trioxide. Varying the concentrations of the reactants is a means of forming a higher concentration of one of the polythionates than the others. Tetrathionates are most easily prepared by oxidation of thiosulfates with iodine. Polythionates are known to be found in Wackenroder's liquid, a solution which contains colloidal sulfur, and which is p[illegible] pared by passing hydrogen sulfide through sul[illegible] rous acid solution.

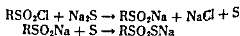
Persulfuric acids. The persulfuric acids (pero[illegible] monosulfuric acid, called Caro's acid, and pero[illegible] disulfuric acid, called Marshall's acid), are known as the acids and salts, and are usually prepared by the electrolysis of sulfuric acid (or sulfate) so[illegible] lutions or by treatment of chlorosulfonic acid (or salt) solutions with hydrogen peroxide The mono- acid is a hygroscopic crystalline material which melts at 45°C. It is soluble in water, alcohol, ether, and organic acids. One proton is readily lost in water, and the other is strongly held The di[illegible] acid is a hygroscopic crystalline substance which melts at 65°C with decomposition It is hydrolyzed by water to the mono- acid and sulfuric acid, and eventually to hydrogen peroxide and oxygen It is also a powerful oxidant, and both acids (and salt solutions) liberate iodine from iodides readily Both acids are used to oxidize organic materials, as bleaching agents, and in the preparation of hydrogen peroxide.

Sulfenic acids. Sulfenic acids are known as the esters and halides. The ethyl ester (C_2H_5—S—O—C_2H_5) can be prepared from C_2H_5S—C[illegible] and sodium ethylate at 0°C. It is a colorless liquid with a foul odor which boils at 108°C (724 mm Hg). It is a weak reducing agent and is oxidized by ethyl hypochlorite to the sulfinic ester, C_2H_5—S(O)(OC_2H_5).

Sulfinic acids. Sulfinic acids are formed by the reduction of the chlorides of sulfonic acids with zinc or by the reaction of Grignard reagents on sulfur dioxide in ether solution. They are unstable in air and are chlorinated by thionyl chloride to give their own acid chlorides, R—S(O)(Cl). Esters of sulfinic acids (for example, C_2H_5—S(O)(OC_2H_5), bp 60°C at 18 mm Hg) are prepared from ester chlorides of sulfurous acid, R—O—S(O)(Cl), and Grignard reagents.

Sulfonic acids. Sulfonic acids (alkyl) are prepared by oxidizing mercaptans (RSH) or alkyl sulfides with concentrated nitric acid, by treatment of sulfites with alkyl halides, or by the oxidation of sulfinic acids. The aromatic derivates (for example, benzenesulfonic acid, $C_6H_5SO_3H$, bp 171°C at 0 1 mm Hg) are prepared by treatment of aromatic hydrocarbons with oleum They are stable substances which are usually water-soluble and can be converted into esters, halides, and amides (which have important medicinal properties). Organic materials are frequently sulfonated in order to render them water-soluble.

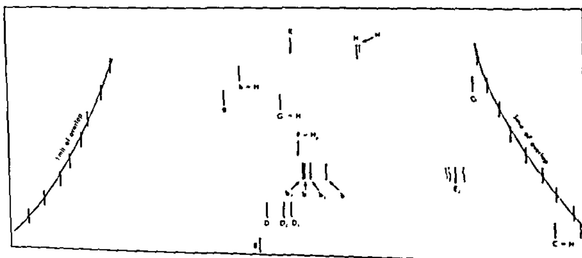
Thiosulfonic acids. Thiosulfonic acids are known as salts and esters. The salts may be obtained from the chlorides of sulfonic acids and sulfides as follows:



$$RSO_2Cl + Na_2S \rightarrow RSO_2Na + NaCl + S$$
$$RSO_2Na + S \rightarrow RSO_2SNa$$

The salts react with alkyl iodides to form the esters, RSO_2SR.

Spectrum of solar flare (above) and identification of key lines (below). Wavelength increases from top down and from right to left; spectrum is spread into successive strips with increasing overlap toward the shorter wavelengths. The right half of the lowest strip is obscured by oxygen absorption in the terrestrial atmosphere. The D_3 line of helium is visible only in flares, chromosphere, and prominences. The $H\alpha$, $H\beta$, $H\gamma$, $H\delta$, and $H\epsilon$ are lines of the Balmer series of hydrogen. *(Sacramento Peak Observatory)*



Line	Wavelength in Angstroms	Element
B	6870	oxygen
$H\alpha$ = C	6563	hydrogen
D_1	5896	sodium
D	5890	sodium
D	5876	helium
E	5270	iron
b_1	5184	magnesium
b	5173	magnesium
b	5167.5	iron

Line	Wavelength in Angstroms	Element
b_1	5167.3	magnesium
$H\beta$ = F	4861	hydrogen
$H\gamma$ = G	4340	hydrogen
G	4308	iron and calcium
g	4227	calcium
$H\delta$ = h	4102	hydrogen
$H\epsilon$	3971	hydrogen
H	3968	calcium
K	3934	calcium

The concentrated acid acts as a strong oxidizing agent because of its tendency to lose an atom of oxygen to form sulfurous acid (H_2SO_3), which readily decomposes to sulfur dioxide (SO_2) and water. Concentrated sulfuric acid reacts with most metals upon heating to produce sulfur dioxide,



$$2Ag + H_2SO_4 \rightarrow Ag_2SO_4 + SO_2 + 2H_2O$$

Gold reacts least readily.

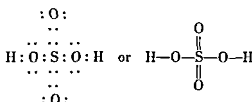
The concentrated acid decomposes salts of other lower boiling acids:



$$H_2SO_4 + NaCl \rightarrow HCl + NaHSO_4$$

It is therefore widely used in the preparation of other acids.

Sulfuric acid ionizes in water, forming hydrogen (H^+), bisulfate (HSO_4^-), and sulfate (SO_4^{--}) ions. The structural formula of sulfuric acid is usually written as



This structure is but one of several in which the molecule exists. It is therefore a resonance hybrid. *See* SULFATE; SULFUR. [F.J.J.]

Sun

The star at the center of the solar system, and the principal source of light and heat for the Earth (Fig 1). The Sun consists of a globe of gas, 1.4×10^6 km in diameter, heated to incandescence by thermonuclear reactions in its deep interior. It is a typical member of the most numerous class of stars, those of spectral type dG2, with surface temperatures of around 6000°K. Other characteristics are given in Table 1.

Fig. 1. Sun, photographed in white light during 1957 maximum of sunspot cycle. (*Mount Wilson and Palomar Observatories*)

SOLAR STRUCTURE

The only star near enough for detailed study, the Sun is of immense astronomical importance as a prototype for stars throughout the universe (*see* STAR). The light and heat from the Sun make the Earth habitable for organic life (*see* EARTH; INSOLATION). Less obviously, the Sun is the ultimate source of nearly all the energy utilized by industrial civilizations, in the form of water power, fuels and wind. (Atomic energy, radioactivity, and the lunar tides are examples of nonsolar energy.)

Solar atmosphere. The main body of the Sun is opaque and has a rather sharply defined visible surface known as the photosphere, the source from which practically all the light and heat of the Sun is radiated. Resting on the photosphere and visible to a height of about 10,000 km is the chromosphere, a complicated layer of transparent tenuous gas shot through with a fur of small luminous spikes known as spicules. Above the chromosphere is the transparent corona, a rather irregular faint appendage of extremely low density and high temperature, extending outward to a distance of several solar diameters. Both the chromosphere and the corona emit a steady flux of radio waves, which have been observed over the frequency range from 15 to 30,000 Mc.

The photosphere, chromosphere, and corona constitute the solar atmosphere and are the only parts of the Sun accessible to observation. They all have the same chemical composition, which probably differs little from that of the solar interior. They are permanent features, always in a state of internal change but always present at all points on the solar surface.

Within the solar atmosphere a number of isolated temporary phenomena occur, much as clouds, thunderstorms, and tornadoes occur in the terres-

Table 1. Principal physical characteristics of the Sun

Mean distance from Earth (the astronomical unit)	$(1\ 4960 \pm\ 0003) \times 10^8$ km
Radius	$(6\ 960 \pm\ 001) \times 10^5$ km
Mass	$(1\ 991 \pm\ 002) \times 10^{33}$ g
Mean density	$1\ 410 \pm\ 002$ g/cm^3
Surface gravity	$(2\ 738 \pm\ 003) \times 10^4$ cm/sec^2 = 28 × terrestrial gravity
Total energy output	$(3\ 86 \pm\ 03) \times 10^{33}$ erg/sec
Energy flux at surface	$(6\ 34 \pm\ 07) \times 10^{10}$ erg/(cm^2)(sec)
Effective surface temperature	5780° ± 50°K
Stellar magnitude (photovisual)	−26 73 ± 03
Absolute magnitude (photovisual)	+4 84 ± 03
Inclination of axis of rotation to ecliptic	7°
Period of rotation	About 27 days The Sun does not rotate as a solid body; it exhibits a systematic increase in period from 25 days at the equator to 31 days at the poles

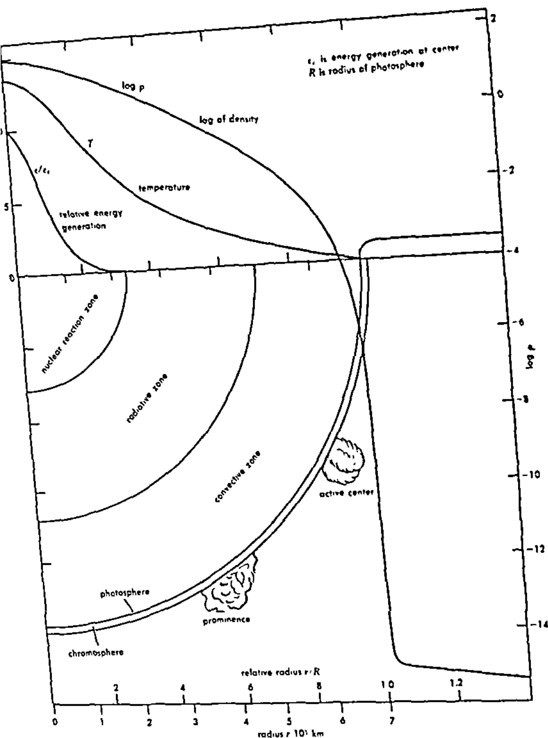


Fig 2 Solar temperature, density, and energy generation in the interior and atmosphere of Sun The curves in the upper section show these as functions of radial distance from the center The lower section shows the principal zones in the interior of Sun.

trial atmosphere. Known collectively as solar activity, they include sunspots and faculae in the photosphere, flares, plages, and spicules in the chromosphere, and prominences and a variety of changing coronal structures in coronal space. Some of the chromospheric and coronal phenomena are · h sudden bursts of radio emission, which often exceed the steady background emission by factors of hundreds or thousands in the lower frequencies.

The material of the Sun is a gas of neutral and ionized atoms, free electrons, and a barely detectable trace of a few of the hardiest molecules. The high temperature is sufficient to vaporize the most

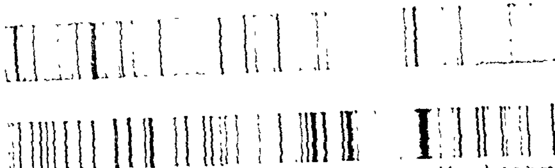


Fig 3 Two sections of the Fraunhofer spectrum, showing bright continuum and dark absorption lines. The wavelength range covered by each strip is approximately 85 A. (Sacramento Peak Observatory)

Surface temperature A knowledge of the total solar radiation and its wavelength distribution permits several estimates of the *surface temperature* of the Sun, all of which depend upon the assumption that the Sun is a black-body radiator (see BLACK BODY). Assuming this, the energy distribution and total radiation from 1 cm² of the solar surface can be accurately represented by Planck's law and Stefan's law

$$E_\lambda = \frac{2hc^2}{\lambda^5}\frac{1}{e^{hc/\lambda kT} - 1}$$

and

$$E = \sigma T^4$$

where c = velocity of light, 2.998×10^{10} cm/sec
h = Planck's constant, 6.62×10^{-27} erg-sec
k = Boltzmann's constant, 1.38×10^{-16} erg/deg
σ = Stefan's constant, 5.672×10^{-5} erg/(cm²)(deg⁴)(sec)

The solar constant gives $E = 6.346 \times 10^{10}$ erg/(cm²)(sec) at the solar surface, from which the effective temperature is found to be $T_e = 5780°$K, by Stefan's law

A comparison of the energy distribution with E_λ calculated from Planck's law reveals notable discrepancies. A color temperature of around 6000°K gives the best over-all fit in the 2900–25,000 A region of the spectrum. Other determinations of temperature from the energy distribution in shorter spectral regions yield a variety of color temperatures between 5500° and 7000°K. The differences between these color temperatures and the effective temperature are simply indications that the Sun is not a perfect black body, under the circumstances there is no unique quantity which can be termed the temperature. The deviation from black-body radiation laws is partly due to the fact that the radiation at different wavelengths comes from layers at different depths in the solar atmosphere with different temperatures.

Solar physics. The combination of a telescope and spectrograph has proved to be the most powerful observational tool for the study of solar physics. See ASTRONOMICAL SPECTROSCOPY.

The solar spectrum is the classical example of an absorption spectrum, consisting of a bright continuum interrupted by thousands of dark absorption lines, known as the Fraunhofer lines. Of the 92 natural elements, 63 are represented in the Fraunhofer spectrum (Fig. 3). The remaining atoms are undoubtedly present but remain undetected because of low abundance or inaccessibility of lines in the far ultraviolet, where the spectrum is absorbed by the Earth's atmosphere. The relative abundances of the most numerous atoms have been estimated from the line intensities. They are listed in Table 2. The Sun can be described as a globe of hydrogen and helium with traces of the other elements.

Table 2. Relative numbers of the most abundant atoms in the Sun

Hydrogen, H	1,000,000		Silicon, Si	20
Helium, He	50,000	200,000	Sulfur, S	8
Oxygen, O	500		Aluminum, Al	2
Nitrogen, N	400		Sodium, Na	2
Carbon, C	200		Calcium, Ca	1.5
Magnesium, Mg	33		Iron, Fe	1.5

Kirchhoff's classical laws of spectrum analysis gave the first sound concept of the structure of the solar atmosphere. Later studies of spectroscopic details and the application of the methods of quantum mechanics have elaborated and extended the picture without essential change. The photosphere is the source of both the continuum and the weaker Fraunhofer lines. The strong lines, because of their opacity, originate at higher levels in the lower chromosphere.

The development of quantum mechanics in the 1920s introduced a period of rapid advance in the interpretation of the solar spectrum. The influence of excitation and ionization on line intensities, and their dependence on the temperature and density of an atmosphere, were clarified. The effects of kinetic temperature, turbulence, density, numbers of atoms in the line of sight, and electric fields on line profiles were recognized, although the exact quantitative relations are still undetermined in some areas. The well known Doppler and Zeeman effects provided means for measuring the line of sight components of velocity and magnetic field strength. The analysis is complicated because all the physical parameters vary with height, and the profile of every line is the composite result of all of them over a considerable range of height. However, different lines are affected differently, and

refractory substances, and shattering collisions reduce molecules to their single atom constituents.

The sunspots and faculae are readily observed with a small telescope, by projection of the solar image through the eyepiece onto a shaded white card. (This is the only safe method for observing the Sun without a specially designed solar eyepiece.) Observations of the chromosphere and corona require additional spectroscopic accessories.

Internal structure. The interior of the Sun can be studied only by calculation. In principle, the problem is to determine the radial distributions of temperature $T(r)$, density $\rho(r)$, and energy generation $\epsilon(r)$, required to maintain a stable gaseous sphere of the Sun's mass, radius, and luminosity (Fig. 2). The solution is complicated by uncertainties in the nature of the interaction of matter and radiation under the extremes of temperature and density of the solar interior. Different assumptions have led to a number of theoretical models, all of which agree on the broad characteristics of the solar interior, although they differ in detail. The temperature at the Sun's center is near 14,000,000°K and decreases gradually outward to 5000°K in the photosphere. The central density is about 90 g/cm³, and decreases to 10^{-7} g/cm³ (about 10^{-4} times atmospheric density) in the photosphere. The energy radiated from the photosphere is the ultimate result of nuclear processes, which consume 4.3 metric tons of the Sun's mass each second in the central region of high temperature and density, extending out to a radius of about 150,000 km. *See* NUCLEAR PHYSICS

In the core of the Sun, and out to a radius of about 5×10^5 km, the energy is transported by the radiative transfer of photons, which bounce from atom to atom like balls on a slightly inclined pinboard. Outside this radiative core is the convective zone which extends to within a few thousand kilometers of the photosphere. Here energy is carried by the bouyant upward motion of heated material through a steep temperature gradient, and the downward motion of the cooled material. Thermal conduction is negligible.

Solar radiation. The solar radiation emerging from the Sun presumably contains electromagnetic energy at all wavelengths (*see* ELECTROMAGNETIC RADIATION). Appreciable quantities of radiation have been measured over the range from 30-m radio waves down to the x-ray region near 10 A (the latter from rockets). More than 95% of the energy, however, is concentrated in the relatively narrow band between 2900 and 25,000 A and is accessible to routine observation from ground stations. The total radiation and its distribution according to wavelength within the observable range are parameters of fundamental significance. They are measures of the total energy output of the Sun and its effective surface temperature.

Solar constant. The solar constant relates to total solar radiation. It is defined as the radiation in calories per minute received on an area of 1 cm² normal to the direction of the Sun outside the Earth's atmosphere, when the Earth is at its mean distance from the Sun (1 astronomical unit). It is an exceedingly difficult quantity to determine from below the atmosphere and is still uncertain to more than 4% after a century of measurement. The accepted value is 1.97 cal/(cm²)(min). This is equivalent to 1.374×10^6 erg/(cm²)(sec).

The measurement of the solar constant is greatly complicated by the absorption of solar radiation at all wavelengths by the Earth's atmosphere. The absorption varies enormously with wavelength, and its wavelength distribution varies with time and location on the Earth. Hence a determination of the solar constant involves the measurement of total radiation received at the ground, the relative distribution of that radiation according to wavelength, and a determination of the fraction absorbed by the atmosphere at each wavelength. These data are sufficient to define the solar constant and the spectral distribution of solar radiation outside the terrestrial atmosphere. The corrections for absorption can be calculated from the observable changes in absorption as the length of atmospheric path varies with solar altitude throughout the day, provided the properties of the atmosphere remain constant over the period of observation. Because such stability rarely occurs and there is no certain method for distinguishing when it does, the atmospheric effects can be eliminated only by a statistical analysis of long series of observations, preferably from widely separated stations where the atmospheric changes can be assumed to be independent.

The basic instrument for the absolute measurement of total radiation is the water-flow pyrheliometer, invented by C. G. Abbot. Water flowing at a known rate through a tube in a blackened enclosure can be heated either by solar radiation or by an electric coil of known resistance R. The measurement consists in determining the electric current I in the coil, which has the same heating effect as the solar radiation. The energy of the radiation must then be $0.24I^2R$ cal/sec.

The distribution of radiant energy according to wavelength has been measured by a variety of receivers in the focal plane of a suitable spectrograph. The most useful are bolometers and thermocouples, which detect changes in the temperature of a minute bit of metal. If the metal is properly blackened to absorb all radiations efficiently, its temperature and hence the output signal are nearly independent of the wavelength of the impinging radiation and vary with the energy alone (*see* BOLOMETER, THERMOCOUPLE). These devices give relative radiant energies at different wavelengths, which can be converted into absolute values (calories or ergs per unit wavelength interval per second) by comparison with a pyrheliometer.

Although many investigators have contributed to the study of solar radiation, the great bulk of the work has been performed by the Smithsonian Institution in a continuous program of observation from the 1880s until 1955, under the direction of S. P. Langley, C. G. Abbo[illegible] Alden.





5 Traces of the longitudinal component of mag-
c field strength over the disk of the undisturbed Sun on two successive days. (Mount Wilson and Palomar Observatories)

h a transmission band 3–4 A wide centered on Hα line (see BIREFRINGENCE) The filter al ist completely suppresses the continuous specum of the overpowering background of sunlight attered by the terrestrial atmosphere while eely transmitting the chromospheric light Fig 8)

The limb photographs of the chromosphere give ie impression of a continuous medium which rapdly thins out with increasing height, becoming invisible 6000–7000 km above the photosphere The picules project from this layer like the bristles of a decrepit hair brush Their numbers diminish with increasing height up to a maximum height of about 15 000 km *In only rare instances do their* apparent diameters exceed the 0.5-sec limit of resolution in the best photographs (due to poor seeing) The average true diameter must be something less than 400 km

An individual spicule typically shoots upward from the lower *chromosphere* at 20–30 km sec to its maximum height There it pauses and *either* fades out or retracts into the chromosphere. The average lifetime is about 3 *min*. Above the level of 4000 km the spicules are generally well resolved and easily counted There appear to be about 10^{4} of them over the whole Sun, with a total area of cross section of not more than 2×10^{-4} of the solar surface

Just as the photosphere is a transition layer *in* opacity, the chromosphere is a transition layer *in* temperature In round numbers, the temperature *at* the base is 4300°K (the minimum of the temperature-*height curve*) *and* rises to about 1,000 000°K at 10,000–15,000 km, where the chromosphere merges with the isothermal corona. The corresponding densities in numbers of atoms/cm³ are A = 10^{14} and 4×10^{9}. While these boundary values of temperature and density are fairly well established, their profiles through the chromosphere *are uncertain.*

The greater temperature and lesser density of the chromosphere as compared with the photosphere are evident from a *comparison of their spectra. The* outstanding characteristic of the chromospheric spectrum is a *great enhancement* of high excitation lines and the lines of ionized atoms over the lines of the photospheric spectrum The strong chromospheric lines of He I and *He II*, with *excitation* potentials in excess of 20 electron volts (compared with 4–6 volts for most photospheric lines) are

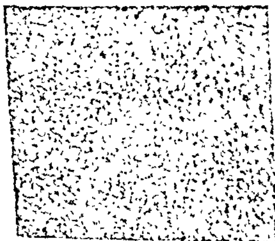


Fig 6. Large-scale photograph of photospheric granulation *in* white *light* taken from an altitude of 80,000 ft above sea level. *The* length of this section is about 35,000 *km* on the Sun (*Princeton Observatory*)

judicious studies of many lines have yielded most present knowledge of solar physics.

Photosphere. The photosphere has been described as the visible surface of the Sun. Actually it is a layer several hundred kilometers thick, the *distinguishing mark of which is a rather abrupt* transition from the high opacity of the solar interior to the nearly perfect transparency of the chromosphere. Photons from below the photosphere cannot escape from the Sun directly. Instead, they are absorbed by atoms in higher layers and reemitted. Photons from the top of the photosphere, however, pass unimpeded through the chromosphere and corona to outer space. As seen from Earth, the photosphere is analogous to a cloud layer in the terrestrial atmosphere. Although light filters through, it is thoroughly scattered, and no image of the source can be seen.

The brightness of the photosphere decreases smoothly from the center of the solar disk to the limb This limb darkening results from the fact that the line of sight to the observer passes through the solar atmosphere at an increasing angle to the normal as the point of observation approaches the limb. Hence the line of sight penetrates to a lower depth in the foggy photosphere at the center of the disk than at the limb, where the path from a given level through the overlying layers is much longer. The fact that the photosphere is darker at the limb than at the center indicates at once that the effective temperature decreases with increasing height through the photosphere While this conclusion is qualitatively correct, the measurement of limb darkening in a broad spectral band cannot yield very meaningful results because the radiation at short wavelengths comes from deeper layers than that at long wavelengths. However, the center-to-limb variations in narrow bands of the solar continuum and in the profiles of the Fraunhofer lines are among the most powerful tools for the analysis of the vertical structure of the photosphere In spite of the difficult problems of interpretation, these data provide a reasonably reliable determination of the variation with height in the photosphere of temperature T, density ρ, and radial optical depth τ_0 where τ_0 is defined as $-\ln k$, where k is the fraction of radiation at $\lambda = 5010$ A, which escapes unabsorbed through the overlying layers in the vertical direction (Fig. 4).

Magnetic field. The Zeeman effect of the large magnetic fields of the sunspots has been observed since 1908, but the field of the undisturbed Sun was too small for detection until the brilliant invention of a photoelectric solar magnetograph by Horace and Harold Babcock in 1952. A typical solar magnetic chart shows the longitudinal (line of sight) component of the photospheric fields during a min- ... (Fig. 5). The deviations ... ontal lines meas- ... al between lines ... ure the field ... corresponds to about 1 gauss. Over most of the Sun the fields are random in direction and only a fraction of a gauss in strength, with a few small areas of several gauss. At the north and south poles, however, the fields are systematically of opposite sign, indicating that the Sun, like the Earth, has a general field.

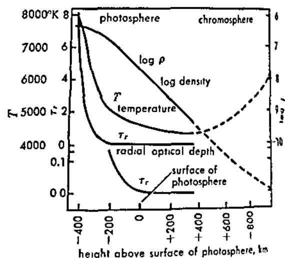


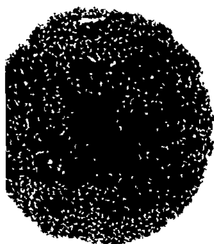
Fig. 4. The variation of temperature, density, and radial optical depth with height in the photosphere and (less certainly) in the chromosphere. The lower curve of optical depth shows the variation with an expanded ordinate scale.

Surface granulation. The only visible structure in the undisturbed photosphere is the white light surface granulation (Fig. 6). It has the general appearance of small white grains like rice sprinkled at random on a gray background The grains are small, of the order of 1000 km in diameter (about 1 3 sec of arc as seen from the Earth), with an average life of about 3 min.

The granulation is probably the visible evidence of convection from below the photosphere The bright grains are presumably the tops of hot rising columns which bring energy up from the interior, while the darker intergranular area is the cooled downward moving material This view is supported by Doppler measurements, which show an apparent over-all upward motion of the photosphere, due presumably to the predominating effect of the rising bright granules with respect to the dark areas The measured difference in brightness between granule and intergranular areas is about 15%, indicating an effective temperature difference of the order of 200°K.

Chromosphere. The chromosphere was first detected and named by early solar eclipse observers They saw it as a beautiful rosy arc that remained visible for a few seconds above the limb of the moon when the photosphere had been covered. The red color is due to the predominating brightness of the Hα line of hydrogen at 6562 8 A in the chromospheric spectrum, which, except for a bare trace of continuum, is a pure emission spectrum of bright lines (Fig. 7).

Because of the rarity of eclipse opportunities solar astronomers have developed methods for photographing the chromosphere with observatory telescopes whenever the sky is clear. The structure at the limb is visible thro ... fringent filte



9 Spectroheliogram of the Sun in Hα light show- the structure of the chromosphere over the disk. unt Wilson and Palomar Observatories)



ig 10 The solar corona photographed by Van iesbroeck at the eclipse of February 25, 1952 (Na- onal Geographic Society)

bright lines by which it can be discriminated, not only from the K and F components, but also from the atmospheric halo that surrounds the un eclipsed Sun

In spite of rather formidable technical difficulties, the coronagraph makes it possible to record the spectrum of the E corona, and even to photograph its structure directly through a birefringent filter, whenever the sky is sufficiently clear (*see* CORONAGRAPH). The spectrograph dilutes the sky light by spreading it into a long continuous spectrum The coronal emission lines, on the other hand, are merely separated by the dispersion without dilution and stand out conspicuously against the sky continuum. Figure 11 shows a short length of spectrum with the green 5303 A line of the corona, and lines of a bright prominence, superposed on the Fraunhofer spectrum of the sky scatter. The advantage of continuous observation, showing activity over periods of hours or days, is, of course, tremendous. Current knowledge of the E corona rests primarily on such observations of the iron green line of Fe XIV at 5303 A, the iron red line of Fe X at 6374 A, and the calcium high temperature yellow line of Ca XV at 5694 A.

The identification of the emission lines was one of the classical astronomical mysteries until B. Edlén solved the problem theoretically in 1942. He showed that they were forbidden lines of highly ionized atoms, which could be excited only under conditions of high temperature, low density, and enormous volumes quite beyond any *conceivable* laboratory resources. Identification is lacking for a few of the 27 known lines, but most of them originate in Fe, Ni, Ca, or Ar from which 9-13 electrons *have* been stripped by the fierce bombardment of neighboring particles. The mere presence of these *ions*, with ionization potentials from 233 to 814 electron volts, is unequivocal evidence of *the high* kinetic temperature of the corona The temperature calculated from the ionization, *which is* determined from the observed absolute and relative line intensities, is about 800 000°K.

The thermal Doppler broadening of the *lines* yields an independent measure of the temperature. If it is assumed that turbulence can be neglected, the line widths *indicate* a temperature of about

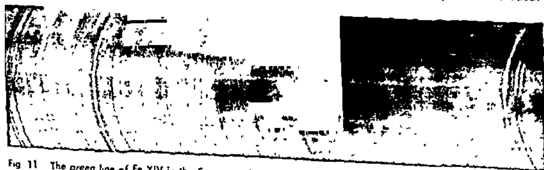


Fig 11 The green line of Fe XIV in the E corona (the long arc at the left) with bright metallic lines of a short prominence, and the yellow chromospheric helium line D_3 (arc at right). (Sacramento *Peak Observatory*)



Fig 7. The flash spectrum of the ultraviolet light from the chromosphere, photographed during an eclipse. This is a negative to enhance details The strong H and K lines are at the right, the convergence of the Balmer series of hydrogen is near the center, and at the left the lines merge into the Balmer continuum toward shorter wavelengths (*High Altitude Observatory*)

entirely absent in the Fraunhofer spectrum. Low density and high temperature favor the high excitation observed.

Like the vertical structure, the horizontal structure of the chromosphere is inhomogeneous (Fig 9). The characteristic small-scale dark mottling indicates definite differences in the physical states of contiguous areas Whether the dark mottles are hotter or cooler than their surroundings cannot be determined until better data on the height dependence of temperature are available

Corona. The corona, which appears in all its glory during total solar eclipses, lies above the chromosphere. It has been universally and correctly described as a pearly white halo surrounding the eclipsed Sun, extending out to one or two solar diameters, with conspicuous streamers reaching far beyond this (Fig 10). The surface brightness of the inner corona is about 10^{-6} times that of the Sun, and the total light emission approximates that of the full Moon

The light originates in three distinct processes, which distinguish the F, K, and E components. The F component is not a true solar appendage It is a halo produced by the scattering of sunlight by interplanetary dust between the Sun and the Earth, and is properly regarded as the inner zone of the zodiacal light (*see* ZODIACAL LIGHT) The F component is negligible compared with the K component in the inner corona, but is brighter than the K component beyond about 2.5 solar radii.

Unlike the F component, the K and E components are intimately associated with the Sun itself The light originates in a tenuous gaseous envelope surrounding the Sun with a maximum density of the order of 4×10^{8} atoms/cm^3, at a temperature of $1\text{–}2 \times 10^{6}$ °K.

At a total solar eclipse, the K corona dominates the picture with its intricate structure, visible in the inner corona, and its beautiful streamers Its luminosity is due entirely to the scattering of sunlight by free electrons in the coronal gas The light is partially polarized. The spectrum consists of a simple continuum with no absorption lines. The absence of the Fraunhofer lines in the K component provided the first clue to the high kinetic temperature of the corona. The Doppler broadening of the lines near 5000 A due to the thermal velocities of the scattering electrons, for which the atomic weight is 5.5×10^{-4}, is roughly $\delta\lambda_T = 0.16\, T^{1/2}$A Thus a temperature of 10^{6} °K or more smears the lines out over 100 A or so and dilutes their absorption to the point of undetectability.

The free electrons of the K corona are the principal source of solar radio emission at frequencies of less than 500 Mc, and may be regarded as a sort of radio photosphere. Measurement of the radio emission indicates a coronal temperature near 10^{6} °K.

The structure of the K corona varies markedly with the sunspot cycle At sunspot maximum it presents a fairly symmetrical globular appearance with a few weak streamers in the equatorial regions. At sunspot minimum the globular shell shrinks toward the solar surface and the streamers emerge The polar brushes composed of small, symmetrically diverging streamers and the long radial streamers in lower latitudes are characteristic.

The E component of coronal light is the only part which originates in the coronal atoms themselves and is accordingly the most informative about the physical state of the corona. Although only one-hundredth as bright as the K component, the E corona has a pure emission spectrum of

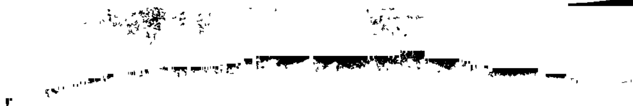
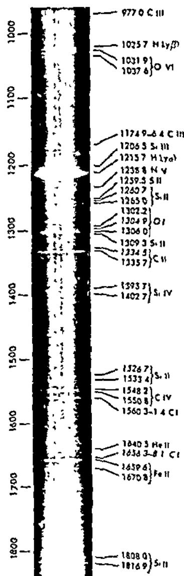


Fig 8 Large-scale photograph of the chromosphere in Hα light, with the disk of the Sun artificially eclipsed and the hairy spicules projecting above the continuous chromosphere The length of this section is about 140,000 km. (*Photograph by R. B. Dunn using 15-in telescope, Sacramento Peak Observ r*)



g 13 High-altitude spectrum of Sun showing emission lines in the *far ultraviolet* of carbon C, helium He, hydrogen H, iron *Fe*, *nitrogen N*, oxygen O, silicon Si, and sodium Na (*U.S. Naval Research Laboratory*)

ection, since its integrated intensity is certainly less than 10^{-5} times that of the visible spectrum

SOLAR ACTIVITY

The foregoing sections have been concerned with the structure of the normal steady state of the undisturbed Sun There are, *in addition*, a number of temporary phenomena, *continually* changing, often explosively, known collectively as solar activity. Most of them are the products of a single more basic phenomenon known as a center of activity

Center of activity. *A* center of activity (CA) is a limited region of the solar surface, typically 150,000 km in diameter, where a strong magnetic field temporarily appears. It is distinguishable from the undisturbed solar surface by a great variety of subsidiary phenomena which will be described below. The maximum magnetic field strength within a center may be about 50-4000 gauss The active lifetime of different centers varies from 3-4 days for a small CA, to about 300 days for a strong one. The longer lived centers may then degenerate into an expanded weak static field of 1-2 gauss which endures another 200-300 days or until a new CA appears in the area. In terms of *L*, the duration of the period of active magnetic changes, the evolution of the more important centers with $L > 50$ days consists of a growth phase from the first appearance to $0.1L$, a period of maximum field strength from $0.1L$ to $0.2L$, and a phase of declining field strength from $0.2L$ to $1.0L$. The interval from $0.0L$ to 0.2-$0.3L$ is a period visibly and radioactively spectacular. Sunspots, plages, concentrations of the E corona, and a barrage of radio emission develop and decay. Short lived flares, with an imposing retinue of related activities, appear with increasing frequency and then fade away. These phenomena are far more conspicuous than the magnetic field, which is measurable only by the relatively subtle Zeeman splitting of some of the Fraunhofer lines. There is little doubt, however, that the magnetic field is the primary characteristic of a CA; the other features are secondary.

Although the complex interaction between the material of the solar atmosphere and magnetic fields is far from thoroughly understood the energies involved favor the theory that the secondary features are relatively trivial results of minor fluctuations in the CA field. *Observational data* on the evolution of the fields are still insufficient *for a detailed* confirmation of this new concept, but *observations* in 1958 and 1959 support it.

The majority of centers have bipolar *fields with* two strong poles, or maxima of field strength, of *opposite* sign in well-developed sunspots. Centers *with unipolar fields* are less numerous, but not uncommon. A few centers have *complex fields*, with many poles of both signs, dominated by a strong bipolar pair Such centers are by far the most active.

Intensities vary enormously from the large sunspot *groups, with* field strengths up to 4000 gauss over areas of 10^9–10^{10} km², down to weak fields of about 50 gauss, marked by small plages The strong centers which develop *sunspots are* confined to the sunspot zone between ±40° latitude. Weaker centers occur with [illegible]
tuc [illegible]

un [illegible]
th [illegible] changes in magnetic fields must be accompanied by electric currents of *enormous* inductive inertia Once established, a field decays slowly, and no *perceptible* changes would be detectable [illegible]
fore app [illegible]
relativel [illegible]
rial belo [illegible]
[illegible] They are occasionall

1,800,000°K, except for occasional Ca XV hot spots which go much higher. The temperatures derived from either line intensities or line widths vary little with height, and the corona appears to be approximately isothermal. Because the thermal conductivity must be high, this is to be expected.

The discrepancy between the line-width value and the ionization and radio determinations of temperature has not been fully explained. Although there are considerable uncertainties in all three, they hardly seem sufficient to account for the differences.

The high temperatures of the chromosphere and corona present an interesting problem. No simple transport of heat from the solar interior is possible, because the relatively cold photosphere intervenes. Several processes have been suggested. The most promising is the dissipation of upward-moving mechanical energy. The original nature of this energy is uncertain It may be moving material in spicules, or acoustic waves modified by the action of local magnetic fields In either event, the dissipation is probably in the form of hydromagnetic shock waves, by means of which particle velocities in excess of the thermal velocities of protons in the corona can be achieved.

The E corona is far more spotty in its distribution over the surface of the Sun and displays much more complex structures than the K corona (Fig 12). Coronal emission tends to concentrate strongly over active centers in the sunspot zones, although some emission is present over the whole Sun. Rapidly changing structures are fairly common They will be discussed later with other features of solar activity.

Observations from rockets. The scientific use of rockets has contributed information of the utmost importance for the understanding of the Sun. The experiments have been concerned with the study of the solar spectrum at wavelengths of less than 2900 A, where the radiation is completely absorbed in the upper atmosphere of the Earth The technical problems encountered in this work are enormous, and any successful observation at all must be counted a brilliant performance Two types of experiment have been carried out a number of times. The first is the direct photography of the solar spectrum to a short-wave limit which includes the fundamental line of the He II spectrum at 304 A. The second is the detection of the total radiation in rather broad spectral bands from about 1300 A down into the x-ray region at 5 A.

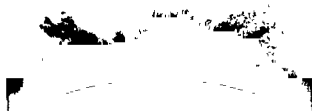


Fig. 12. Typical stable form of structure of the E corona in the light of the green line of Fe XIV photographed without an eclipse through a birefringent filter. (Sacramento Peak Observatory)

The two major problems in the spectroscopic work are the difficulty of constructing an optical system which is at the same time sufficiently transparent and sufficiently sophisticated to obtain the desired data, and the requirement that the instrument be accurately pointed at the Sun from a rapidly and erratically moving platform Successful results have been achieved, however, by W. Rense of the University of Colorado, A. Jursa of the Geophysics Research Directorate, U.S Air Force, and by a group at the U.S. Naval Research Laboratory under R. Tousey. The work of these experimenters has resulted in preliminary knowledge of the principal features of the solar spectrum over the whole range down to 300 A

The spectrum of closely spaced absorption lines extends down to about 1600 A, where, for some unknown reason, the continuum suddenly decreases in intensity to below the detectable level At shorter wavelengths, the spectra show nearly 50 bright emission lines, many of which originate in such highly ionized atoms as O VI, N V, C IV, and Si IV (Fig. 13). The presence of so many emission lines was a surprise They probably originate in the transition region of intermediate temperature between the lower chromosphere and corona The most important results, however, were the detection of the extremely bright fundamental Lyman alpha line of hydrogen at 1216 A, and the corresponding line of He II at 304 A. The presence of these lines was predicted by theory, but needed decisive confirmation. An accurate measurement of their intensities and profiles is a difficult problem for the future, but one of the utmost importance in the study of the chromosphere These resonance lines are much more simply related to the temperature and density in the chromosphere than are the subsidiary lines in the visible spectrum, which depend more on the population of the excited quantum levels by the radiation field.

The measurement of x-rays below 50 A and a comparison of their intensities with the Lyman alpha line have been performed at the Naval Research Laboratory under H. Friedman and T. Chubb The radiations were detected in ionization chambers with windows of various materials which served as filters The results indicate that appreciable radiations in these wavelength regions are always emitted by the Sun. The intensities of the x-rays are enormously enhanced during the occurrence of a flare. Although the data are somewhat uncertain, it appears that the corresponding enhancement of Lyman alpha is small.

It is fortunate that at wavelengths shorter than 1600 A the solar radiation is largely concentrated into emission lines which are far more easily photographed than a continuum Otherwise, the far ultraviolet spectrum of the Sun might have defied de-

Fig 13. High-altitude spectrum of Sun showing emission lines in the far ultraviolet of carbon C, helium He, hydrogen H, iron Fe, nitrogen N, oxygen O, silicon Si, and sodium Na. (U S Naval Research Laboratory)

tection, since its integrated intensity is certainly less than 10^{-5} times that of the visible spectrum.

SOLAR ACTIVITY

The foregoing sections have been concerned with the structure of the normal steady state of the undisturbed Sun. There are, in addition, a number of temporary phenomena, continually changing, often explosively, known collectively as solar activity. Most of them are the products of a single more basic phenomenon known as a center of activity.

Center of activity. A center of activity (CA) is a limited region of the solar surface, typically 150,000 km in diameter, where a strong magnetic field temporarily appears. It is distinguishable from the undisturbed solar surface by a great variety of subsidiary phenomena which will be described below. The maximum magnetic field strength within a center may be about 50–4000 gauss. The active lifetime of different centers varies from 3–4 days for a small CA, to about 300 days for a strong one. The longer-lived centers may then degenerate into an expanded weak static field of 1–2 gauss which endures another 200–300 days or until a new CA appears in the area. In terms of L, the duration of the period of active magnetic changes, the evolution of the more important centers with $L > 50$ days consists of a growth phase from the first appearance to $0.1L$, a period of maximum field strength from $0.1L$ to $0.2L$, and a phase of declining field strength from $0.2L$ to $1.0L$. The interval from $0.0L$ to 0.2–$0.3L$ is a period visibly and radioactively spectacular. Sunspots, plages, concentrations of the E corona, and a barrage of radio emission develop and decay. Short lived flares, with an imposing retinue of related activities, appear with increasing frequency and then fade away. These phenomena are far more conspicuous than the magnetic field, which is measurable only by the relatively subtle Zeeman splitting of some of the Fraunhofer lines. There is little doubt, however, that the magnetic field is the primary characteristic of a CA; the other features are secondary.

Although the complex interaction between the material of the solar atmosphere and magnetic fields is far from thoroughly understood, the energies involved favor the theory that the secondary features are relatively trivial results of minor fluctuations in the CA field. Observational data on the evolution of the fields are still insufficient for a detailed confirmation of this new concept, but observations in 1958 and 1959 support it.

The majority of centers have bipolar fields with two strong poles, or maxima of field strength, of opposite sign in well-developed sunspots. Centers with unipolar fields are less numerous, but not uncommon. A few centers have complex fields, with many poles of both signs, dominated by a strong bipolar pair. Such centers are by far the most active.

Intensities vary enormously from the large sunspot groups, with field strengths up to 4000 gauss over areas of 10^9–10^{10} km², down to weak fields of about 50 gauss, marked by small plages. The strong centers which develop sunspots are confined to the sunspot zone between ±40° latitude. Weaker centers occur with decreasing frequency at high latitudes, but are rarely found above 60°.

[illegible]

un [illegible]

th[illegible]

be [illegible] by electric currents of enormous inductive inertia. Once established, a field decays slowly, and no perceptible changes would be detectable over [illegible]

fore appea [illegible]

relatively p[illegible]

rial below [illegible] atmosphere. They are occasionally

1,800,000°K, except for occasional Ca XV hot spots which go much higher. The temperatures derived from either line intensities or line widths vary little with height, and the corona appears to be approximately isothermal. Because the thermal conductivity must be high, this is to be expected.

The discrepancy between the line-width value and the ionization and radio determinations of temperature has not been fully explained Although there are considerable uncertainties in all three, they hardly seem sufficient to account for the differences.

The high temperatures of the chromosphere and corona present an interesting problem. No simple transport of heat from the solar interior is possible, because the relatively cold photosphere intervenes Several processes have been suggested. The most promising is the dissipation of upward-moving mechanical energy. The original nature of this energy is uncertain. It may be moving material in spicules, or acoustic waves modified by the action of local magnetic fields. In either event, the dissipation is probably in the form of hydromagnetic shock waves, by means of which particle velocities *in excess of the thermal velocities of protons in the* corona can be achieved

The E corona is far more spotty in its distribution over the surface of the Sun and displays much more complex structures than the K corona (Fig 12). Coronal emission tends to concentrate strongly over active centers in the sunspot zones, although some emission is present over the whole Sun Rapidly changing structures are fairly common They will be discussed later with other features of solar activity.

Observations from rockets. The scientific use of rockets has contributed information of the utmost importance for the understanding of the Sun The experiments have been concerned with the study of the solar spectrum at wavelengths of less than 2900 A, where the radiation is completely absorbed in the upper atmosphere of the Earth. The technical problems encountered in this work are enormous, and any successful observation at all must be counted a brilliant performance. Two types of experiment have been carried out a number of times. The first is the direct photography of the solar spectrum to a short-wave limit which includes the fundamental line of the He II spectrum at 304 A. The second is the detection of the total radiation in rather broad spectral bands from about 1300 A down into the x-ray region at 5 A.

The two major problems in the spectroscopic work are the difficulty of constructing an optical system which is at the same time sufficiently transparent and sufficiently sophisticated to obtain the desired data, and the requirement that the instrument be accurately pointed at the Sun from a rapidly and erratically moving platform Successful results have been achieved, however, by W. Rense of the University of Colorado, A. Jursa of the Geophysics Research Directorate, U.S Air Force, and by a group at the U.S. Naval Research Laboratory under R. Tousey. The work of these experimenters has resulted in preliminary knowledge of the principal features of the solar spectrum over the whole range down to 300 A.

The spectrum of closely spaced absorption lines extends down to about 1600 A, where, for some unknown reason, the continuum suddenly decreases in intensity to below the detectable level At shorter wavelengths, the spectra *show nearly* 50 bright emission lines, many of which originate in such highly ionized atoms as O VI, N V, C IV, and Si IV (Fig 13). The presence of so many emission lines was a surprise. They probably originate in the transition region of intermediate temperature between the lower chromosphere and corona The most important results, however, were the detection of the extremely bright fundamental Lyman alpha line of hydrogen at 1216 A, and the corresponding line of He II at 304 A. The presence of these lines was predicted by theory, but needed decisive confirmation. An accurate measurement of their intensities and profiles is a difficult problem for the future, but one of the utmost importance in the study of the chromosphere. These resonance lines are much more simply related to the temperature and density in the chromosphere than are the subsidiary lines in the visible spectrum, which depend more on the population of the excited quantum levels by the radiation field.

The measurement of x-rays below 50 A and a comparison of their intensities with the Lyman alpha line have been performed at the Naval Research Laboratory under H. Friedman and T Chubb The radiations were detected in ionization chambers with windows of various materials which served as filters. The results indicate that appreciable radiations in these wavelength regions are always emitted by the Sun The intensities of the x-rays are enormously enhanced during the occurrence of a flare Although the data are somewhat uncertain, it appears that the corresponding enhancement of Lyman alpha is small.

It is fortunate that at wavelengths shorter than 1600 A the solar radiation is largely concentrated into emission lines which are far more easily photographed than a continuum Otherwise, the far ultraviolet spectrum of the Sun might have defied de-

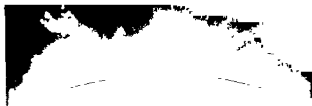


Fig. 12. Typical stable form of structure of the E corona in the light of the green line of Fe XIV photographed without an eclipse through a birefringent filter. (Sacramento Peak Observatory)



Fig 15. Characteristic small prominence, showing fibrous structure. The generally vertical filaments are paths of downward moving material. Horizontal width of a single frame is 125,000 km. Hα light (Sacramento Peak Observatory)

not generally appear to be associated with the centers of activity in which they originate, although the low latitude end may *terminate in*, or *cut across*, a center

One of the most surprising phenomena of the large filaments is a tendency to disappear suddenly for a few days and then to reappear in the original place and form. At least some of these sudden disappearances are due to eruptions like that shown at the limb in Fig. 14. The tendency for a prominence to reappear with little change suggests that its existence is due to some rather stable condition of the surface below, which survives the eruptive cataclysm

Another characteristic phenomenon is the tendency of material from large prominences to flow downward into the chromosphere. There is usually a continuous rain of detached knots, and often material flows in a steady stream along one or more fixed arcs, which start horizontally and curve down into a single center of attraction. The prominence of Fig. 14 shows such a streamer. Although material flows through them quite rapidly, these streamers are remarkably stationary in quiet prominences like rigid tracks which may well be the lines of force in a magnetic field. The main body of the prominence appears to be inexhaustible, con- space where conditions are such that downward-moving material becomes luminous as it passes through. The corona is the most likely source for this material.

Spectra of prominences Prominences of all types are usually somewhat fainter than the chromosphere. Their spectra are similar. The three spectra of Fig. 16 are typical of bright quiet prominences. The middle spectrum, with low dispersion, includes the strong H and K lines and extends toward shorter wavelengths beyond the Balmer limit at 3647 A. The upper and lower photographs show different sections of this region with higher dispersion. The distortions of the lines are due to the Doppler effect of motions within the prominences. All three exposures show a faint but appreciable prominence continuum due to scatter by free electrons. The background of Fraunhofer spectrum is due to light scattered in the terrestrial atmosphere.

The chemical composition of the prominences is presumably identical with that of the chromosphere and photosphere. The density is about 10^{10} or 10^{11} hydrogen atoms/cm^3. The kinetic temperatures vary widely. The large hedgerows are the coolest, at 7000°–10,000°K. The more active types are hotter, with temperatures which may be of the order of 10^5 °K.

Active prominences and the remaining features of solar activity are more closely associated with



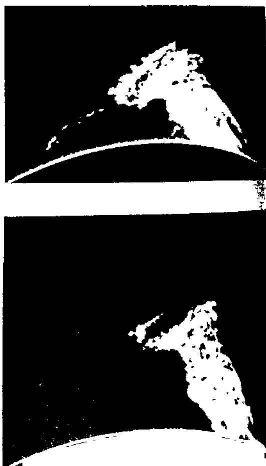


Fig 14. Large prominence in eruptive phase after days or weeks of static inactivity The four frames cover an interval of 29 min. Horizontal width of single frame is about 670,000 km. Hα light (Sacrament Peak Observatory)

brought to the visible surface in a CA by an unknown mechanism, possibly the subphotospheric convection.

The connection of the different features of solar activity with centers of activity is least evident in the large stable prominences, although this independence is more apparent than real. The prominences will therefore be considered first, followed by a description of the phenomena more closely related to the centers of activity.

Solar prominences. The most beautiful appendages of the Sun, and perhaps the strangest, are the solar prominences. They appear in Hα light at the limb as great red clouds of gas, sometimes resting on the surface of the Sun, but frequently floating free with no visible connection (Figs. 14 and 15). Most prominences at the limb are at least a thousand times fainter than the photosphere, but, like the chromosphere and corona, they have an emission spectrum by means of which they can be discriminated from the scattered light of the sky with the aid of a birefringent filter. Against the solar disk they strongly absorb Hα and the H and K lines of Ca II, appearing as long dark filaments

The prominences display a variety of shapes, sizes, and activities that defy any general description. Their common characteristic is that they appear above the chromosphere as bodies of gas at 100–1000 times the density and at a much lower temperature than the material of the surrounding corona, with which they may well be in pressur equilibrium The fibrous structure is typical The show little evidence of influence by solar gravit and must be supported by forces far more powe ful, almost certainly magnetic fields of a few gau or more. Here the common features end. A fe characteristic types of prominence can be distu guished, and indeed, there have been a number worthy efforts to classify them. The most signi cant criteria are probably the association, or lac of it, with a CA, and the predominance of upwar moving material from the chromosphere or downward-moving material from the corona Th origin and driving forces activating the differe classes are diverse, and their inclusion under th common term prominence is somewhat misleadin

Hedgerow prominence. Prominences of the lar est class, known as quiescent or hedgerow pro nences, are identical with the long, dark filamen on the solar disk (Fig 14). A typical filament h the form of a thin sheet, a few thousand kilomete thick and about 200,000 km long, standing ver cally above the chromosphere to a height of 400 km. The average lifetime is about 100 days Th filament grows from a short, dark exclamatio mark (!) extending poleward from the perio which is represented by a CA. As it grows, it ther drifts away or is simply drawn out under th influence of the latitude shear of solar rotatio Thus the fully developed hedgerow prominence





Fig 17 Centers of activity photographed in Hα light The length of each frame is about 280,000 km on Sun. (Sacramento Peak Observatory)

sharp, having the appearance of discontinuities in brightness A small spot may consist only of an umbra; the large spots in bipolar groups are generally more complicated, with several separated umbrae in a single penumbral area.

A sunspot begins as a small dark area known as a pore, 2000–3000 km in diameter. The pore develops into a full-fledged spot in a few days, and the maximum development is reached within the next week or two. Decay, which consists simply of shrinkage of a spot area along with its magnetic field, is much slower The life span varies from a few days for small spots to about 100 days for large groups (Fig 19)

The largest individual spots are about 120,000 km in diameter Spot groups may attain a total length of 250,000 km in the east-west direction.

Activity cycle. One of the most remarkable characteristics of solar activity is a strong cyclic variation in its intensity, with a period of about 11 years The cycle was first discovered in the sunspots simply because they are the most easily observed features of solar activity, and it was named the sunspot cycle before the other phenomena were recognized Although the sunspot cycle is evident in all forms of solar activity (and for some the effect is even more pronounced than in the sunspots), the term sunspot cycle will be used here.

As manifested by the sunspots, the cycle consists of variations in the sizes and numbers of spots, expressed quantitatively as the sunspot number (Fig 20) The cycle also involves changes in sunspot heliocentric latitudes Each cycle begins with a few spots in high latitude between 20° and 35° in the northern and southern hemispheres. As the cycle progresses the spots increase in number and size until they reach a maximum in about 4 years The decay takes about twice as long, and the cycle ends with a few small spots between latitudes of 5° and 10° The progression from high to low latitude is quite steady throughout the cycle. Oddly, the equator appears to be a zone of avoidance where few spots are found The characteristic disposition of spots on the solar disk at the maximum of 1957 is shown in Fig. 1.

The latitude behavior of sunspots suggests that the commencement of each cycle is a new beginning, rather than a revival of decaying activity The high latitude spots of a new cycle usually appear before the low latitude spots of the preceding cycle are gone, and successive cycles overlap considerably Most of the spots develop in the magnetic fields of bipolar centers of activity, which exhibit a peculiar systematic behavior in sign Throughout the life of a cycle the field of the leading spot is almost without exception of one sign in the northern hemisphere and the opposite sign in the southern hemisphere The sign of the field in the following spot is always opposite to that of the leading spot. More remarkable still, all signs reverse in successive cycles.

Since the maximum of 1750, the average interval between successive maxima has been 10.9 years

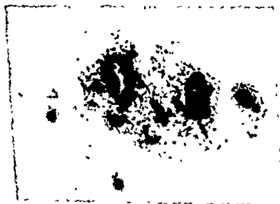


Fig 18 Large sunspot group of May 17, 1951, photographed in white light shows the filamentary structure of the penumbra and the granulation of the surrounding surface (Mount Wilson and Palomar Observatories)

Fig 16. Spectra of bright prominences at different dispersions with wavelengths of corresponding lines. From top to bottom the wavelength range covered by each spectrum is 88,545, and 129 A. The converging Balmer series merges into the Balmer continuum in bottom spectrum. (*Sacramento Peak Observatory*)

centers of activity than the hedgerow prominences or filaments The activity is complex (Fig. 17). The phenomena fall naturally into two fairly distinct classes, the equable features and the eruptive features. The equable features are present in every CA and evolve gradually during the active phase of the center. They are the first and last visible evidences of its existence, although the magnetic field persists long after their disappearance. The eruptive phenomena, on the other hand, change rapidly, enduring for only a few minutes or hours The majority of the centers are primarily equable, with little eruptive activity Those with complex magnetic fields are more explosive, with a continual display of eruptive phenomena during the 10-20 days of greatest activity.

The principal equable features are sunspots, faculae, plages, enhanced coronal regions, and low-frequency radio noise. The eruptive features include loop and coronal prominences, rapid changes in the E corona, and flares. The latter are often accompanied by surge prominences, coronal hot spots, radio sources which apparently shoot up through the corona, bursts of ultraviolet radiation and x-rays, corpuscular showers, and occasionally a great increase in cosmic ray emission

Sunspots. The sunspots are the most conspicuous features of a CA. They are dark areas on the Sun, squarely centered on the strongest magnetic fields They sometimes occur singly, in unipolar centers More often they appear in a complicated group dominated by two large spots in a bipolar center

Sunspots are darker than the surrounding photosphere (about 18% as bright), simply because they are cooler The refrigerating mechanism is not definitely known, but the vertical magnetic field is probably responsible In an area of great field strength the normal convection which maintains the temperature of the photosphere should be inhibited The field permits motion of material only along the lines of force (*see* MAGNETOHYDRODYNAMICS) Thus, any lateral adiabatic expansion, and the lateral transfer of material between upward and downward moving convection columns is prevented, and convection ceases. Deprived of its normal supply of heat, the area cools by radiation and becomes a dark sunspot

The structure of a typical sunspot appears as a dark central area, the umbra, surrounded by a gray ring known as the penumbra (Fig 18). The inner and outer boundaries of the penumbra are





Fig 17 Centers of activity photographed in Hα light. The length of each frame is about 280,000 km on Sun (*Sacramento Peak Observatory*)

sharp, having the appearance of discontinuities in brightness. A small spot may consist only of an umbra; the large spots in bipolar groups are generally more complicated, with several separated umbrae in a single penumbral area

A sunspot begins as a small dark area known as a pore, 2000–3000 km in diameter The pore develops into a full fledged spot in a few days, and maximum development is reached within the t week or two Decay, which consists simply of :inkage of a spot area along with its magnetic ld is much slower. The life span varies from a w days for small spots to about 100 days for large oups (Fig 19).

The largest individual spots are about 120,000 n in diameter Spot groups may attain a total ngth of 250,000 km in the east-west direction.

Activity cycle. One of the most remarkable characteristics of solar activity is a strong cyclic variation in its intensity, with a period of about 11 *years* he cycle was first discovered in the sunspots imply because they are the most easily observed eatures of solar activity, and it was named the *unspot* cycle before the other phenomena were recognized. Although the sunspot cycle is evident in all forms of solar activity (and for some the effect is even more pronounced than in the sunspots), the term sunspot cycle will be used here.

As manifested by the sunspots, the cycle consists of variations in the sizes and numbers of spots, expressed quantitatively as the sunspot number (Fig 20). The cycle also involves changes in sunspot heliocentric latitudes Each cycle begins with a few spots ... in tl ... cycl ... size ... reach a maximum in about 4 years The decay takes about twice as long, and the cycle ends with a few small spots between latitudes of 5° and 10°. The progression from high to low latitude is quite steady throughout the cycle Oddly, the equator appears to be a zone of avoidance where few spots are found The characteristic disposition of spots on the solar disk at the maximum of 1957 is shown in Fig 1.

The latitude behavior of sunspots suggests that the commencement of each cycle is a new beginning, rather than a revival of decaying activity. The high latitude spots of a new cycle usually appear before the low latitude *spots of the preceding* cycle are gone, and successive cycles overlap considerably. Most of the spots develop in the magnetic fields of bipolar centers of activity, which exhibit a peculiar systematic behavior in sign. Throughout the life of a cycle the field of the leading spot is almost without exception of one sign in the *northern* hemisphere and the opposite sign in the southern hemisphere. The sign of the field in the following spot is always opposite to that of the leading spot. More remarkable still, all signs reverse in successive cycles.

Since the maximum of 1750, the average interval between successive maxima has been 10 9 years.

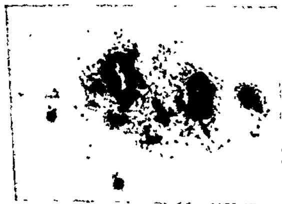


Fig 18. Large sunspot group of May 17, 1951, photographed in white light shows the filamentory structure of the penumbra and the granulation of the surrounding surface. (*Mount Wilson and Palomar Observatories*)

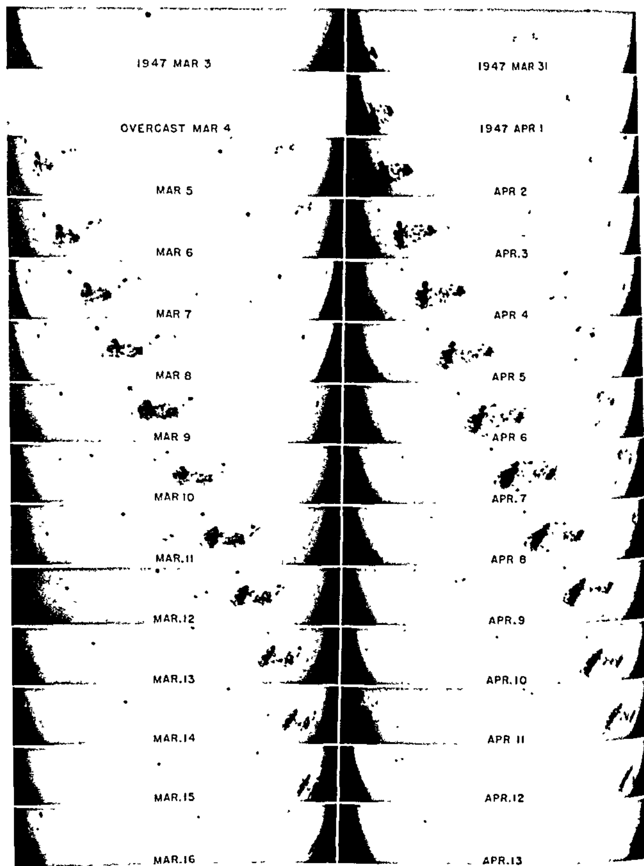


Fig. 19. Development of the large 1947 sunspot group of unusually long life. Successive frames show section of Sun parallel to equator on successive days as group rotated across visible hemisphere. *(Mount Wilson and Palomar Observatories)*

The average duration of a cycle from the first appearance of high latitude spots to the disappearance of the last low latitude spots is nearly 14 years. The difference between these two figures is the result of overlap of consecutive cycles. While the variations in sunspot activity are pronounced,

Fig 20 Annual means of sunspot number, 1750–1956

and the periodicity of the cycle clearly recognizable, the interval between successive maxima is quite irregular, varying from 8 to 17 years.

There is no satisfactory explanation for the sunspot cycle in terms of physical activity in the Sun at present.

Faculae. Sunspots near the limb of the sun are always surrounded by *faculae*. They are small, irregular, bright patches, about 10% brighter than the surrounding photosphere. As a spot rotates toward the center of the disk the photosphere becomes brighter, and the faculae fade into invisibility. They are apparently the white-light appearance of the plages, visible at the limb because of their height above the haziness of the upper photosphere.

Plages. The plages, also known as flocculi or chromospheric faculae, are large, irregular, bright patches surrounding sunspot groups (Fig. 21). They can be seen only with the aid of a spectroheliograph or birefringent filter transmitting either the Hα line of hydrogen or the H or K line of ionized calcium. When photographed under conditions of excellent seeing, plages display an extraordinarily intricate structure in association with small dark filaments, which are frequently arrayed around the sunspots in a systematic radial or spiral pattern (The Hα filtergrams of a CA in Fig 17 show typical patterns.) The plages seen in the calcium lines are usually larger and more conspicuous than in Hα light. Because the calcium lines probably originate lower in the chromosphere than Hα lines, they show the cross section of a plage nearer its base.

Although sunspots are always accompanied by plages the reverse is not true. Occasional fields of small plages develop in the sunspot zone and fade out without the appearance of any spots. The few studies of photospheric magnetic fields in these



Fig. 21. The disk of the Sun photographed in Hα light, showing bright plages in centers of activity and dark filaments (prominences) (Sacramento Peak Observatory)

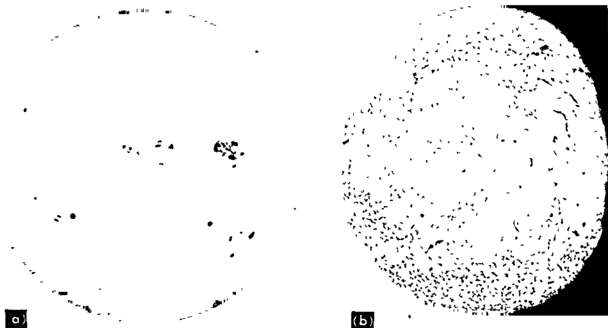


Fig 22. The Sun photographed nearly simultaneously (a) in white light and (b) in red Hα light (*Mount Wilson and Palomar Observatories*)

areas, however, show weak CA of 50–200 gauss field strength

Plages are features of the lower chromosphere and often completely or partially obscure an underlying sunspot, as shown by comparing an Hα spectroheliogram and white light photograph of the Sun taken simultaneously (Fig 22). They usually appear in a center a day or two before the sunspots as small bright areas, which rapidly grow to cover most of the CA When the sunspots have decayed, the plages usually remain for a few weeks to mark the location and are among the last of the active center phenomena to disappear The disappearance is a process of breaking up into a network of bright filaments which gradually fade out.

E coronas. The E coronas show a strong correlation with the plages, developing large regions of enhanced emission in the green and red lines, which progress through nearly the same evolution of growth and decline as the plages, whether accompanied by sunspots or not. The coronal structure often takes the form of the broad static rays diverging from a plage area in the chromosphere (Fig 12). Over a strong CA, a corona exhibits rapid changes of an eruptive character in association with other eruptive phenomena, particularly the flares and loop prominences.

Flares. The flares with their attendant subsidiary phenomena are the most spectacular of the eruptive activities of a CA (Fig. 23) As seen in the Hα line, they are brilliant flecks of light which suddenly appear in a center, almost as though turned on by a switch (Fig 24). The light curve normally consists of a sharp rise to peak brightness in a few minutes, followed by a slower decline of 15 min or more. Areas of flares are comparable to those of sunspots. Large flares are rare, but the numbers increase rapidly for flares of decreasing area, until, at the smallest detectable size, flarelike brightenings are always in progress during the active phase of a center Small brightenings less than 10^4 km in diameter are classed as subflares.

Flare shapes vary systematically with size, from small compact elliptical objects to large coarsely filamentary irregular structures

A flare is usually a chromospheric event, with the limb appearance of a low broad-based dome Internal velocities seem to be low, and the flare phenomenon appears to be simply a sudden brightening of chromospheric material in place. A few notable ex

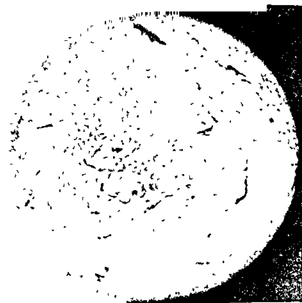


Fig. 23. Unusually large flare of September 18, 1958, photographed in Hα light and enlarged from a 16-mm image. (*Sacramento Peak Observatory*)



Fig 24 Development of a large, coarsely filamentary flare in Hα light Intervals between exposures are 4 min, 19 min, and 13 min. Each frame shows 280,000 km on the Sun (Sacramento Peak Observatory)

ceptions have been observed, however, in the form of flare prominences or sprays (Fig 25) The flare takes the form of a violent ejection of material which tends to break up into small high-speed knots Occasionally a stationary flare appears high in a prominence mass

The most notable features of the flare spectrum are an enormous brightening and broadening of the Hα line and the appearance of normally absent high-excitation lines of helium, either in absorption or emission During the brightest phase of a large flare, many of the Fraunhofer lines are appreciably filled in and weakened, and all the lines of hydrogen, the H and K lines of Ca II, and a few other metallic lines are present in emission

The flare productivity varies widely from one CA to another and appears to be a function of the complexity of the magnetic field Large centers of great activity may have as many as 60-80 flares of which one or two may exceed 80,000 km in the longest dimension. The period of greatest flare productivity is during the week or 10 days when the associated sunspot group goes through its maximum development. According to R. Giovanelli, the rate of flare production is roughly proportional to the product of the sunspot area and its rate of change, whether increasing or decreasing.

Surges. About 80% of the large flares and 30% of the smallest ones are accompanied by surges. A surge is a curved sword of prominence material, thrusting up into the corona and retracting into the Sun along the same path (Fig 26). This decidedly nongravitational behavior is in accord with the concept of material constrained by a fixed magnetic field, moving in the one permitted degree of freedom along the lines of force, like beads on a

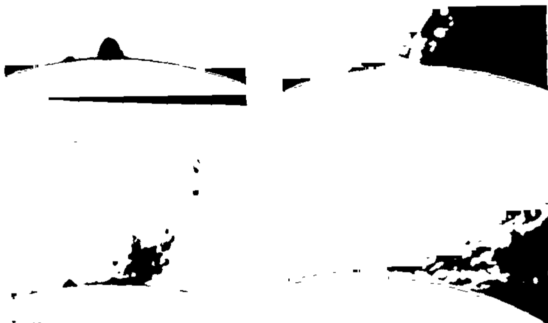


Fig 25 Flare prominence or spray of February 10, 1956 Universal time for successive frames 211600, 211845, 212115, 213345 cover a total interval of 17 min 45 sec. Some fragments exceeded a velocity of 1100 km/sec Each frame shows 670,000 km on the Sun (Sacramento Peak Observatory)

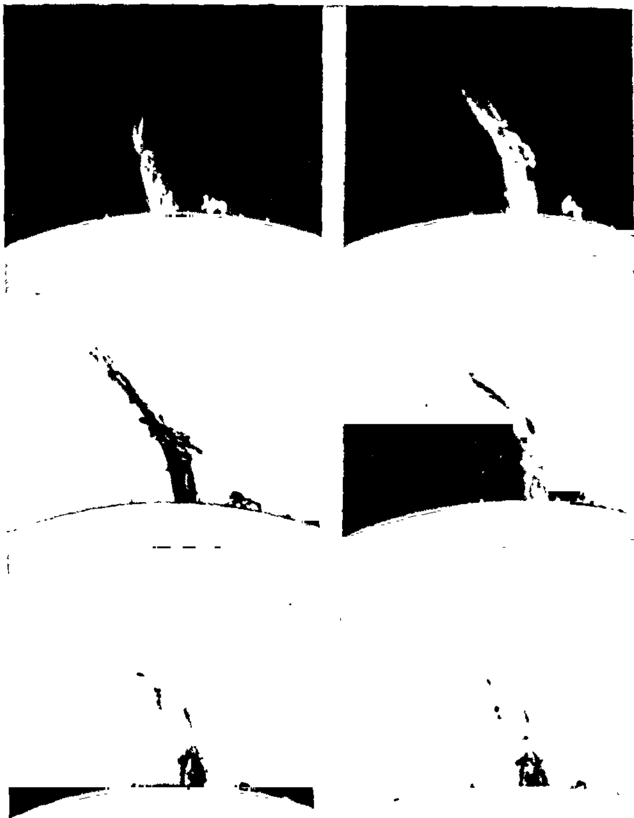


Fig. 26 Photographs show large surge in Hα light during total time interval of 45 min. Frame spans 500,000 km on the Sun. (*Sacramento Peak Observatory*)

wire. Velocities of 300 km/sec and extended lengths up to 150,000 km are typical.

A surge usually appears in an area near to, but not in, a flare. Its trajectory is invariably directed away from the center of a nearby major sunspot. Some surges are apparently unrelated to flares, although these are usually associated with subflares.

Coronal hot spots Coronal hot spots, indicated by emission of the yellow line of Ca XV at 5694 Å above the photosphere by some 30,000 km, have been found in all spectrographic observations of the corona over a limb flare. The yellow line is not normally detectable in the coronal spectrum and indicates a temperature probably in excess of

4,000,000°K. A hot spot is usually visible before the brightening of the flare and lasts longer.

Radio phenomena. Outbursts of radio noise often accompany flares, particularly the large ones. The emission originates in the corona at heights between 50,000 and 500,000 km over the flare. The most interesting observations of radio phenomena are those which sweep the frequency spectrum from 40 Mc to 600 Mc. The corona radiates a given frequency from a fixed height which is characteristic of that particular frequency. Thus a sweep through the radio spectrum is actually a sweep in height through the corona from about 50,000 km at 600 Mc to 700,000 km at 40 Mc. Four distinct flare-associated phenomena have been observed, none of which have been identified with optical events in the corona (Fig. 27).

The Type II burst is a sudden onset of radio noise in a narrow band at high frequency, which *drifts* downward in frequency to less than 100 Mc during an interval of 2–5 min. Type III burst is similar, except that the whole phenomenon occurs within a few seconds. These events are attributed to disturbances in the corona moving upward with velocities of approximately 1,200 km/sec and 100,000 km/sec, respectively.

The U bursts are similar to the preceding, except that the frequency reaches a minimum somewhere above 100 Mc and then rises again to the original frequency.

The fourth kind of flare-associated radio disturbance is the sudden onset of noise with a continuous spectrum over a band of several hundred megacycles, shown in the bottom strip of Fig. 27. It begins during the visible phases of a flare and may last anywhere from 10 min to 5 hours.

Terrestrial disturbances. Sudden disturbances in the terrestrial ionosphere coincident with flares, and geomagnetic storms some hours later, are evidence of the emission of ultraviolet radiation, x-rays, and electrically charged particles at the time of a flare. The sudden ionospheric disturbances due to ultraviolet or x rays are without exception associated with flares, and the most intense magnetic storms are strongly correlated with flares. However, some of the largest flares near the center of the solar disk produce no terrestrial effects. No observable *peculiarities* have been found to distinguish these from the flares associated with the most intense geophysical disturbances.

Relation of flares to CA. To summarize the flare activity, some large flares occur simultaneously



Fig. 27. Time variations in the solar radio spectrum. Frequency range *from* 100 *to* 580 Mc (the vertical coordinate) is divided into *three overlapping* bands with frequencies indicated in top row. Variations with time are shown in *horizontal* direction. Interval between the white pips between *the* middle and lower frequency band [illegible] op left, *a* highly magnified U burst in the 140-Mc band. The interval between successive vertical scans is 0.3 sec. *Top right,* a series of Type III bursts. The middle and lower rows are continuous, *showing* the development of a Type II burst followed by *a strong* continuum in the 450-Mc band. (*Recordings from radio spectrometer, Fort Davis Station, Harvard College Observatory*)

with all of the following subsidiary phenomena: surges, coronal hot spots, radio-emitting disturbances in the high corona, ultraviolet and x-ray emission, and bursts of charged corpuscles. Most sizable flares are associated with at least several of these. Apparently all flares have their coronal hot spots. Of the subsidiary phenomena, only the ultraviolet and x-ray emissions occur exclusively in association with flares. The existence of such a complex of associated activities, some of which occur in widely separated locations within or above a center of activity, strongly reinforces the assumption of some massive underlying driving force, presumably the magnetic field

The remaining eruptive phenomena of a CA, the loop and coronal prominences, and the activity in the E corona mentioned earlier, seem to be independent of individual flares. Loop prominences, however, are the almost infallible mark of a highly eruptive center in its most active phase. When a loop appears at the east limb, flares can be expected during the next 2 weeks as the CA rotates across the solar disk.

Loop prominence. There are several typical forms of loop prominence, all of which are often displayed during the course of a single performance (Fig. 28). The distinguishing characteristic is a small nucleus of prominence material, perhaps 5000 km in diameter, in a fixed position high above the solar surface. From it an apparently inexhaustible supply of material streams down into the chromosphere along one or two sharply defined arcs of circles, the bottoms of which may be cut off by the chromosphere. A single loop may appear by itself, or a number of loops originating in several nuclei may pour material down into the same general area of the chromosphere, giving the appearance of a spiral spring with its loops tied together at the bottom.



Fig. 28. Two characteristic loop prominences in Hα light. Width of Sun in (a) is 340,000 km and in (b) is 125,000 km. (Sacramento Peak Observatory)

The nucleus is commonly located at the center of a coronal hot spot in which the Ca XV lines attain their greatest observed intensities, and may be accompanied by a coronal continuum The downward streaming arcs often radiate strongly in the green coronal line in addition to the usual prominence lines of hydrogen, helium, and the metals. This behavior is shared by a few of the surges. It indicates that loops and surges involve the coronal material at temperatures ranging from 4,000,000°K down to 1,000,000°K, as well as the relatively high-density prominence material at temperatures probably below 100,000°K. The relation between coronal and prominence material is intimate, because many of the details of the prominences photographed in red Hα light are reproduced in the green line. However, the Doppler shift of the green line is always much less than that of Hα. This suggests that the prominence material flows through a sheath of condensed coronal material like water through a hose Possibly the prominences are formed by a continuous condensation of coronal material at the top If so, some unusual process simultaneously compresses and cools coronal material.

Coronal prominences. Loop and coronal prominences are probably related, although coronal prominences are unaccompanied by coronal emission They consist of swarms of small isolated knots of prominence material, elongated in the direction of motion, which cascade down into a single, rather sharply defined area of the chromosphere along converging curved trajectories. A given trajectory will often be well delineated by several of these knots moving in train along it simultaneously. The knots have the appearance of packets of material which are too faint to be visible at the highest levels but become brighter as they move down toward the solar surface They are always in full motion when first detected. The whole phenomenon has the general appearance of a flock of sea birds diving into a school of fish

The corona, as shown by birefringent filter photographs in the green line, is normally in a state of constant change over a CA Coronal loops are a striking example associated with a complex loop prominence (Fig 29). Other changes consist in the development and decay of bright streamerlike structures, and surprisingly dark circular holes in the surrounding material One very characteristic structure is a series of five or six concentric arches over an active center These arches generally expand slowly; a few examples have been observed to break open at the top and whip up into vertical streamers with apparent velocities up to 600 km/sec. This activity seems to coincide accurately with the appearance of a flare.



Fig 29 Six exposures of coronal activity over an interval of 4 hours taken in the green line (5303) Concentric arch structure in first frame is a characteristic structure (Sacramento Peak Observatory)

Solar-terrestrial effects. The direct influences of the Sun on the Earth other than heating, illumination, and the solar tides are generally too delicate for direct detection by human senses, except for the aurora Ultraviolet and x-radiation at wavelengths shorter than 1300 A, and streams of charged particles emitted by the Sun are the active agents The Sun is strongly variable in these radiations, and the induced geophysical effects are pronounced when observed with appropriate radio equipment and refined versions of the magnetic compass

The steady flux of solar ultraviolet and x-radiation maintains the terrestrial ionosphere by ionizing a small fraction of the molecules of the Earth's atmosphere above the 100-km level *See* IONOSPHERE

At the onset of a solar flare the Sun emits a burst of short wave radiation which causes a sudden ionospheric disturbance (SID) The radiation penetrates the upper ionosphere and produces a new layer of ionization at a height of only 60 km Here the density of the air is sufficient to inhibit the oscillation of free electrons by collisions and to absorb radio waves which are normally reflected from higher ionospheric layers (*see* RADIO-WAVE PROPAGATION) Long-distance radio communication deteriorates or is blacked out altogether for a few hours Temporary ionospheric currents produce changes in the geomagnetic field strength, and if severe, they may induce currents in long land lines sufficient to stop wire communications. Every SID is associated with a flare, although a few large flares fail to excite a SID

The wavelength of the solar radiation which produces a SID has not yet been firmly established. The rocket observations of the Naval Research Laboratory, however, indicate that the solar x-ray spectrum is much more strongly enhanced during a flare than any other short-wave region Because the x-rays have sufficient energy to produce the ionization, they are the most likely candidate.

The first solar terrestrial effects to be recognized, in the 1860s, were the high correlations between the 11-year sunspot cycle, the frequency of geomagnetic storms, and the appearances of auroras in abnormally low latitudes, such as those visible in the United States (*see* AURORA; GEOMAGNETISM). At the time, the physical connection was utterly mystifying More careful magnetic observations over a world wide network of stations fully confirm these remarkable correlations.

A geomagnetic storm is a small, rapidly changing perturbation of the relatively massive, steady magnetic field of the Earth It is now known to be induced by the impingement of electrically charged particles (mostly protons and electrons) from the Sun on the permanent geomagnetic field As the particles approach the Earth the permanent field deflects them into trajectories, most of which miss the Earth altogether. A fraction of the particles, however, do penetrate the field and shower into the

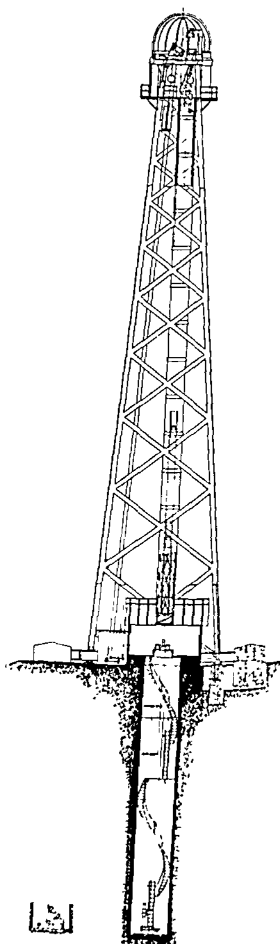


Fig 30 The 150-ft solar tower telescope of the Mount Wilson Observatory has a vertical underground spectrograph.

upper atmosphere in circular zones, normally 25° from the geomagnetic poles. Collisions with atmospheric molecules excite the beautiful luminescence of the aurora, which reaches its maximum intensity in the lower ionosphere at a height of 100 km. The interaction between the incoming particles and the permanent field results in ionospheric currents which induce the small perturbations known as magnetic storms.

Whenever the numbers and velocities of the incoming particles are greatly enhanced, the magnetic disturbances naturally increase, and the auroral zone moves farther away from the geomagnetic pole.

The parallelism between sunspot activity and magnetic storms is due entirely to the dominating effect of the very intense great magnetic storms excited by the corpuscular bursts originating in a CA. Although large flares usually produce the most vigorous geomagnetic disturbances, great magnetic storms often occur without a flare. Whatever the particle-emitting disturbance in a CA may be (possibly a slight fluctuation in its magnetic field), it apparently may or may not produce a flare. However, the eruptive centers are more highly correlated with geomagnetic activity than the quieter centers, even though the magnetic storms may not be associated with individual flares.

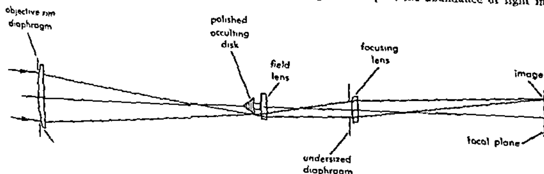
Magnetic storms associated with the very large flares tend to appear about 20 hours after the flare, although intervals from 5 to 70 hours have been recorded. This lag is presumably the time of flight for the solar particles in their passage from the Sun to Earth. The average velocity of the fastest particles in a given burst is usually about 2000 km/sec. Because the large magnetic storms often last for 2 or 3 days, either the particles in a single short-lived burst have a large dispersion in veloc[illegible]

[illegible] consists almost entirely of series of small recurrent magnetic storms. A series begins with a weak disturbance, followed by storms of increasing strength at 27-day intervals until a maximum is reached, after which successive storms become weaker until the series finally dies out. A given series may consist of anywhere from 5 to 20 recurrent disturbances, and several overlapping series are often in progress at once. The particles responsible for these recurrent storms apparently originate in a limited area on the surface of the Sun, known as an M region, which is aimed at the Earth like a gun for a short period during each 27-day rotation of the Sun. The M region presumably emits particles continuously during the whole life of a series of recurrent magnetic storms. No detectable feature of solar activity with the same cycle of development and decay as the M-region storms has been found. Statistical studies show that the M region is a quite undistinguished portion of the solar surface which crosses the central meridian 3–4 days before the corresponding geomagnetic disturbance. This assumption is supported by the fact that the appearance of a sunspot in such a region usually signals the abrupt termination of a series of M storms. Although the evidence is not yet conclusive, C. W. Allen has suggested very plausibly that the K corona streamers are actually the streams of particles from the M regions, which produce the recurrent storms as they sweep across the Earth.

The M storms are clear-cut and easily observed near minima of the sunspot cycle simply because there is little other solar activity at these times to confuse the picture. The M regions may continue to occur [illegible] maxim[illegible] region [illegible] terminates the activity. At sunspot maximum, the much stronger effects of the ever-present centers of activity tend to submerge the recurrent storms and are responsible for the over-all strongly positive correlation between solar and geomagnetic activity.

SOLAR INSTRUMENTS

The observational instruments of the solar astronomer are not basically different from those in general astronomical use. However, a given instrument receives about 10^{3} times as much light from 1 sec² of the solar surface as it does from a tenth-magnitude star. This enormous difference in available light leads to instruments of appreciably different form. Solar telescopes and spectrographs are designed to exploit the abundance of light in



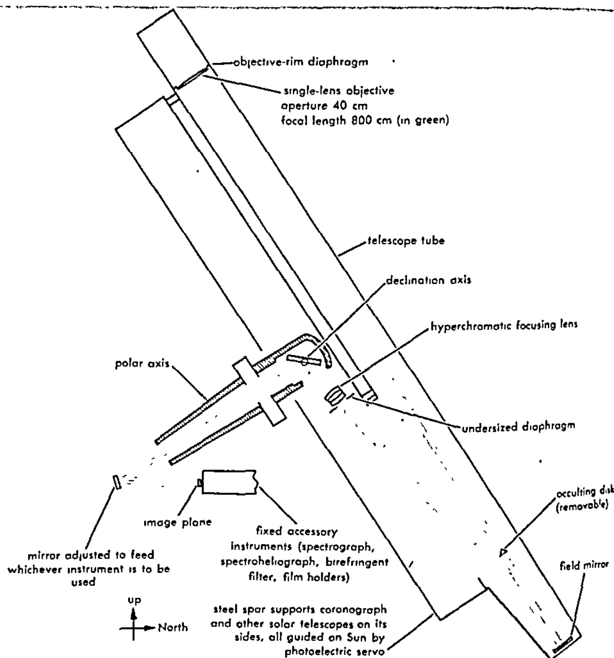


Fig 32. Large 40-cm solar telescope at the Sacramento Peak Observatory.

terms of definition of fine detail and the highest possible spectroscopic power.

Telescopes. Solar telescopes fall into two general classes, those designed for observations of the brilliant solar disk, and coronagraphs designed for the study of the much fainter prominences and still fainter corona through the relatively bright, scattered light of the sky.

Disk telescope. Excellent definition in a solar image of reasonable size is usually the first requirement in a disk telescope. Definition is limited by the excellence of the optical system, by the fundamental diffraction limit of resolution inherent in a given aperture, and by the quality of the seeing (*see* TELESCOPE, ASTRONOMICAL). Seeing is the term used to describe the blurring effect of the changing density gradients in the terrestrial atmosphere due to thermal convection. Seeing is never perfect, and the best that can be hoped for is a location where excellent definition is possible during perhaps 100 hours in a year.

The most convenient type of disk telescope is tl solar tower (Fig 30). Two flat mirrors at the t of a tower, one of which is equatorially mounted follow the diurnal motion of the Sun, reflect su light into a long-focus fixed vertical telescope. Th telescope may be either a refractor or a compou reflector of 30 50 cm aperture. It produces a lar image of the Sun near ground level, where the lig can be conveniently reflected into any one of a num ber of large fixed accessory instruments.

A less expensive variant of the tower telescope the horizontal fixed telescope, with the flat mirro at ground level This arrangement, when careful designed, works well for small apertures, but co vective disturbances in the long horizontal air pa degrade the definition of large instruments.

For some purposes an equatorially mounted tel scope of standard length is preferable to the fix types The elimination of a pair of flat mirrors in proves the definition and considerably reduces sca tered light The telescope m[illegible]ed with a se

ondary optical system to magnify the solar image. Some of the finest photographs of photospheric detail have been taken with such an arrangement on telescopes of 12–20 cm aperture. It is also effective for motion picture observations of prominences and the chromosphere through a birefringent filter, because the filter requires a secondary optical system.

Coronagraph. The coronagraph, invented by B Lyot in 1931, is designed with one overriding consideration in mind: the elimination of instrumental scattered light. Its most delicate task is the observation of the ...

the objective, diffraction at the edge of the aperture, and otherwise insignificant defects in the glass of the objective are all sources of diffuse scattered light which is usually some hundreds of times brighter than the corona.

The coronagraph is deceptively simple (Fig. 31). The critical component is the single lens objective; it must be made of flawless glass which is polished and cleaned to a perfection far beyond ordinary standards. The rest of the system is more straightforward. The disk of the Sun is eclipsed by a polished conical disk, and photospheric light diffracted by the rim of the objective is intercepted in its image formed by the field lens, on the undersized diaphragm. The final lens then images the corona on the focal ...

be ...

at ...

green line may be taken by inserting the appropriate birefringent filter behind the focusing lens and letting the image fall on a photographic film or plate.

The usual coronagraph has an aperture of 15 cm or less. The largest is a 40-cm instrument (Fig. 32). An internal mirror system reflects the light into an observing laboratory through the polar axis, where the image can be directed to large analyzing instruments, as in a tower telescope. The focusing lens corrects the color aberration of the objective, and the telescope can be used equally well for disk and limb observations. (Figures 17 and 24 were taken with it in combination with a birefringent filter.) *See* CORONAGRAPH

Spectrographic accessories. Most modern solar research requires more than the unassisted telescope, the use of which is largely restricted to studies of solar granulation and sunspot detail. Usually the telescope serves to form an image of the Sun for analysis with powerful accessory instruments, the most important of which are the spectrograph, spectroheliograph, and birefringent filter.

Spectrograph. Modern solar spectrographs for use with disk telescopes utilize the great brightness of the solar image to achieve dispersion of the order of 10 mm/A and spectroscopic resolution in excess of 500 000 (*see* SPECTROPHOTOMETRIC ANALYSIS). These desirable characteristics require long focal lengths, 10–25 m, and superlative diffraction gratings of the largest sizes consistent with accuracy of ruling.

Large solar spectrographs are nearly always either of the Littrow autocollimating type, or the reflecting type in which the collimator and camera element are long-focus concave mirrors (Fig. 33). They may be mounted either vertically in a well or horizontally on solid concrete piers.

The optical path inside a large spectrograph is long, particularly in the reflecting type, and internal poor seeing produced by convection in the light path is always a problem. One bold solution certain of success is complete elimination of the air by placing the spectrograph in a tube which can be evacuated. The McMath-Hulbert Observatory has been very successful with such a vacuum spectrograph. Their solar spectrograms are probably the finest that have been taken.

The spectrographs used with coronagraphs for observations of solar limb phenomena are generally much smaller and simpler, with medium dispersion

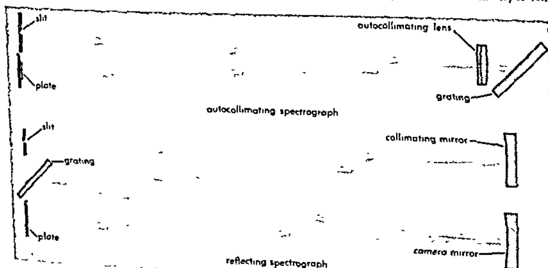


Fig. 33. Two ...

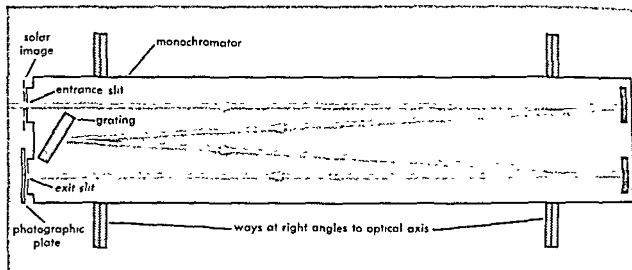


Fig. 34. A spectroheliograph

of 0.5mm/A or less. They are usually simple Littrow grating spectrographs of about 200-cm focal length. The most important requirements are high light efficiency and a stigmatic image. Most of them are small enough to be carried on the same mounting with the coronagraphs that feed them.

Spectroheliograph. The spectroheliograph and birefringent filter present an extended area of the solar image at one sharply defined wavelength. The wavelength chosen is usually at the center of the Hα line of hydrogen or the H or K line of ionized calcium The chromosphere is opaque at these particular wavelengths, and the picture obtained is, therefore, that of the chromosphere. It is evident from Figs. 10, 17, 21, and 22 that the structure is quite different from that of the photosphere, shown in Figs. 1, 6, and 18.

The spectroheliograph is a stigmatic scanning monochromator made by inserting a slit in the focal plane of a spectrograph, accurately centered on the Hα line (or any other wavelength desired), so that only the light of this line is transmitted Variations in image intensity along the first slit are faithfully reproduced in the exit slit Several different scanning arrangements have been successfully used; one uses fixed optics (Fig. 34). The monochromator is mounted on ways which permit smooth motion in the direction of dispersion The solar image from a fixed telescope falls on the entrance slit of the monochromator. The light of the Hα line emerging from the second slit falls on a stationary photographic plate. As the monochromator moves along its ways, the first slit scans across the solar image, and the exit slit correspondingly scans the photographic plate. *Thus an image of the Sun in the light of the Hα line is built up on the photographic plate continuously* See SPECTROHELIOSCOPE.

Birefringent filter. The birefringent filter consists of a multiple sandwich of alternate layers of polarizing films and plates cut from a birefringent crystal (usually *quartz or calcite*). The assembly transmits the light in a series of sharp, widely spaced wavelength bands One or another of the polarizers absorbs the light of all intervening wavelengths A glass filter is usually sufficient to isolate the desired band and exclude the others. Filters made for observations on the disk of the Sun generally have transmission bands 0 5–1.0 A wide, centered on the Hα line. For observations at the limb, bandwidths up to 10 A are used for prominences in Hα, and bandwidths of about 2 A for the green and red coronal lines The birefringent filter is compact enough to be used with a conventional small telescope but is far less flexible than the bulkier spectroheliograph in choice of band width and wavelength. [J.W.E.]

Bibliography: G. Abetti, *The Sun*, rev. ed., 1957; S. Chapman and J. Bartels, *Geomagnetism*, 2 vols., 1940, M A. Ellison, *The Sun and Its Influence*, ... *Physik* vol.

Sundew

Any plant of the genus *Drosera* (90 species) of the order Sarraceniales. Sundews are small, herba-



The sundew, *Drosera* sp., an insectivorous plant (General Biological Supply House)

eous, insectivorous (insect-eating) plants that grow on all the continents, but mainly in Australia Numerous glandular hairs (tentacles) on the leaf secrete a viscous fluid which traps a visiting insect. The tentacles then bend inward about their victim bringing it into contact with the surface of the leaf where it is digested The proteins of the digested insect supply nitrogen, which otherwise may be unavailable to the plant in the place where it grows. The droplets secreted by the glands on the leaves glitter like dewdrops in the light of the morning sun; hence, the name, sundew See INSECTIVOROUS PLANTS; SARRACENIALES; SECRETORY STRUCTURES, PLANT. [P.D.S.]

Sundial

An instrument for telling time by the Sun It is composed of a style that casts a shadow and a dial plate, which is the surface upon which hour lines are marked and upon which the shadow falls. The style lies parallel to Earth's axis. The construction of the hour lines is based on the assumption that the apparent motion of the Sun is always on the celestial equator

Sundials can be made in any form and on any surface They may be large and stationary, or small and portable They may be made for use in a particular place or made for use anywhere The most widely used form is the horizontal dial that indicates local apparent time (Sun time) Other forms of the sundial indicate local mean time, and standard time (see TIME)

The highest form of sundial construction is found in the heliochronometer which tells standard time with great accuracy Incorporated in its construction is the equation of time and the time difference in longitude between the place where it is to be used and the standard time meridian for that locality This makes possible a sundial that can be read as a clock

The sundial is said to be the oldest scientific instrument to come down to us unchanged The underlying scientific principle of its construction makes it a useful device for educational purposes as well as for timekeeping [R.N.M.]

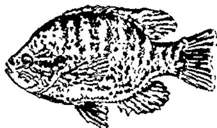
Bibliography: R. N. Mayall and M. L. Mayall *Sundials—How to Know, Use, and Make Them*, reprint, 1958

Sundog

One of two bright spots (or parhelia) which are seen on both sides of the sun or moon, usually with red coloring on the part closer to the sun They are produced by prismatic refraction on the alternate sides of the hexagonal ice crystals floating in the air with vertical axes of symmetry Their angular distances from the sun increase from 22° when the sun is on the horizon to 45° when the sun is at an elevation of 60°. For the sun at the horizon they are situated at the inner halo (see HALO), with higher sun they approach the outer halo. The parhelia are the second most frequently observed halo phenomena. [Z.S.]

Sunfish

A popular name for various fishes. The true sunfishes, family Molidae, are marine fishes with short bodies, extremely short tails, and confluent ventral fins. They can inflate their bodies in the manner of the puffers. Best known is the giant sunfish, *Mola mola*, up to 8 ft long and weighing 1800 lb, which occurs in both the Atlantic and Pacific oceans.



The sunfish, *Lepomis gibbosus*; length to 8 in. (From E. L Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

The fresh-water sunfishes, family Centrarchidae, are North American. They have the spinous and soft parts of the dorsal fin united into a single structure Most of them are laterally compressed, but a few, such as the black basses, are only slightly flattened Sunfishes are frequently called by other names, such as pumpkinseed, bream, bluegill, or, incorrectly, perch. They are prolific and in spite of their small size provide sport fishing. They are excellent food fishes. Best known to fishermen are the black basses, bluegill, and crappies See BASS, CRAPPIE; PERCH, PERCIFORMES. [J.D.B.]

Sunflower

This plant, *Helianthus annuus*, is a native of the Great Plains of North America Varying from 5 to 20 ft in height, it has broad, ovate, and coarsely serrated (toothed) leaves growing from a single stalk See LEAF (BOTANY); STEM (BOTANY) Large showy yellow flowers with dark centers produce flat, round seed heads 10–15 in. across. See FLOWER (BOTANY) Seeds (technically fruits) are about ½ in. long, pointed at one end, and white, black, or striped See FRUIT (BOTANY), SEED (BOTANY)

Distribution and economic importance. Sunflowers are grown in all parts of the United States California, Minnesota, and North Dakota produce most of the domestic, commercial seed crop Production of seed in this country for 1954 was 378,881 bushels, valued at $614 703 on the farm Production in foreign countries, particularly in Russia, far exceeds that of the United States.

In foreign countries, sunflowers are grown mainly for oil Russians eat the seed ...

... used in poultry feed, for birdseed, and as a source of oil The oil is used


Maturing sunflowers on a farm in California (USDA)

Great Plains and Canada, sunflowers are grown for silage for cattle. [A B.B]

Sunflower diseases. Sunflowers are subject to destructive diseases throughout the world

Causal organisms. Rust (*Puccinia helianthi*), caused by an air-borne fungus, reduces seed yields and may kill the plants. Root rot and wilt, caused by a soil-borne fungus, *Sclerotinia sclerotiorum*, affects sunflowers and also many other broad leaved plant species. Other sunflower diseases caused by soil borne fungi include downy mildew, *Plasmopara halstedii*, and leaf mottle, *Verticillium albo-atrum*, which are world-wide in distribution, and black root rot, *Sclerotium bataticola*, which is most common in tropical and subtropical regions *See* FUNGI.

The virus diseases of sunflowers include mosaics and aster yellows (see PLANT VIRUS) They are still somewhat restricted in their geographical distribution.

Control measures. Chemical control of both rust and leaf mottle is possible, but it is not economically feasible. Cultural measures, such as weed removal and rotation with other crops which are disease resistant, help reduce losses from most of the sunflower diseases, and these measures are the only controls available against some of them Sunflowers resistant to rust have been produced in Manitoba, Canada, and varieties resistant to downy mildew, leaf mottle, and aster yellows are being developed there. [W.E.S]

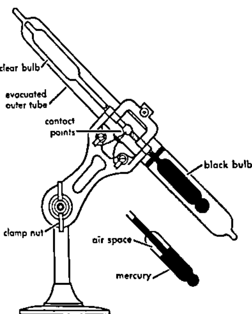
Bibliography: *See* AGRICULTURAL SCIENCE (PLANT); PLANT DISEASE.

Sunlamp

A mercury-vapor gas-discharge tube used to produce ultraviolet radiation for therapeutic or cosmetic purposes. The envelope of the tube is made of quartz, since this is much more transparent to ultraviolet radiation than glass Ultraviolet radiation from a sunlamp extends from visible violet light at about 3800 angstroms (A) to the transmission limit of quartz at about 2800 A. Strong emission at about 2967 A is desirable for treating rickets. Longer wavelengths, approaching the visible spectrum, are more desirable for giving a good suntan in a reasonable time. The wavelength of ultraviolet radiation can be controlled to a great extent by the choice of material for the envelope of the tube, by using special glass filters, and by using fluorescent sunlamps. In fluorescent sunlamps the inside of the long glass tube is coated with a phosphor that gives radiation in the desired range when excited by ionization of the mercury vapor in the tube. *See* ULTRAVIOLET LAMP. [J.M.R]

Sunshine-duration transmitter

A device for transmitting recordable information concerning the number of hours the sun is visible each day compared with the number of hours of possible sunshine These instruments must be sensitive to direct sunshine, but not to sky brightness, even though a cloudy sky near noon might be brighter than direct sunlight near sunrise or sunset The Campbell-Stokes instrument uses a clear glass sphere to focus an image of the sun on a paper card held in a bowl concentric to the glass sphere The sun's image burns a trace in the paper card to provide a record Near sunrise and sunset the image may not be bright enough to burn the paper The Maring-Marvin recorder is a clear bulb and black-bulb differential gas thermometer encased in a clear glass vacuum jacket to isolate it from ambient air. In direct sunlight the warmer black bulb moves a mercury column in a straight connecting tube between the bulbs toward contacts embedded in the tube. The contact closure is used to record the presence of sunshine. Very bright sky near noon on an overcast day may cause this device to indicate a visible sun in error. The above-mentioned defects are overcome in the photoelectric sunshine recorder adopted by the U.S. Weather Bureau In this, two self-generating photoelectric cells are connected in opposition to control a relay. One of the cells is protected from direct sunlight by a shade ring, the otl . ' exposed to both



A sunshine-duration transmitter of the clear-bulb–black bulb or differential thermometer type (From F A Berry, Jr, E Bollay, and N R Beers, *Handbook of Meteorology*, McGraw-Hill, 1945)

sunlight and skylight When direct sunlight is present, the output of the shaded cell is much less and the imbalance closes a relay to actuate the recorder Sunshine recorders usually have a provision to adjust for changes in sun angle as the year progresses *See* METEOROLOGICAL INSTRUMENTATION [V.E.S.]

Sunspot

A dark area in the photosphere of the Sun caused by a lowered surface temperature The temperature at the center of a spot is about 4000°K, and the surface brightness is one-fifth that of the normal photosphere The sizes and numbers of sunspots vary in the celebrated 11-year sunspot cycle, which is shared by all other forms of solar activity *See* SUN. [J.W.E.]

Superaerodynamics

That branch of gas dynamics dealing with the flow of gases at such low density that the molecular mean free path is not negligibly small. Under these conditions the gas no longer behaves as a continuous fluid Important modifications in flow phenomena occur which are ascribable to the discrete molecular structure of the gas The subject is also often called rarefied gas dynamics

Early work in the field was done by the kinetic theory physicists of the nineteenth century This work was mostly concerned with low speed flow of gas at low pressures through orifices and ducts. Following World War II there has been a revived interest in the field due to problems arising from high speed flight at extremely high altitudes This more recent work has thus been more concerned with the supersonic flow of low-pressure gas past aerodynamic objects.

Flow regimes. It is convenient to divide superaerodynamics into three flow regimes These are called free molecule flow, transition flow, and slip flow, corresponding respectively to highly rarefied, moderately rarefied, and only slightly rarefied flow conditions. Phenomena in the three regimes are quite dissimilar, so the subdivision is useful. The term "rarefied" is relative, but can be made quantitative in terms of a dimensionless number, called the Knudsen number K, defined as the ratio of the mean free path λ divided by a characteristic dimension L of the flow field, for example, the diameter of a duct or the length of a high altitude rocket ($K = \lambda/L$). Free molecule flow corresponds to $K \gg 1$, slip flow to $K \ll 1$, and transition to intermediate values of K.

The Knudsen number is related to two other basic parameters of fluid mechanics, namely the Mach number $M = V/a$ and the Reynolds number $Re = VL/\nu$, where V is a characteristic velocity of the flow, a is the speed of sound, and ν the kinematic viscosity. Kinetic theory shows that $\nu \sim a\lambda$; it follows that $K \sim M/Re$, and in particular,

$$K = \sqrt{\frac{\pi}{2}\gamma}\,\frac{M}{Re}$$

where γ is the isentropic exponent.

Free molecule flow. The regime of highly rarefied flow corresponds in the aerodynamic case to flight at altitudes of 100 miles or more The mean free path is long compared to dimension L. It follows that molecules, which impinge on a surface and are then reemitted, travel very far before colliding with another molecule. The essential simplification of free molecule flow is that such collisions can be neglected For the aerodynamic case of flow past a convex body the incident molecules are thus not disturbed by the presence of the body. For many applications it is correct to assume that the distribution of molecular velocities for these incident molecules is in Maxwellian equilibrium but with a mean streaming velocity component

net flux then is determined by the nature of the distribution of velocities for the reemitted molecules. Detailed knowledge of this sort of molecule-surface interaction has not yet been obtained. Under some circumstances the reemitted molecules emerge from the surface in Maxwellian equilibrium with it. Such reflection is called diffuse However, under other circumstances this does not happen. The departure from diffuse reflection depends on the physical and chemical nature of the surface material and of the gas, upon their temperatures, upon the length of time the surface has been exposed to the gas at some particular pressure, and upon the nature of any film of adsorbed gas adhering to the surface For engineering applications it has been customary to define three surface inter-



Maturing sunflowers on a farm in California. (USDA)

Great Plains and Canada, sunflowers are grown for silage for cattle. [A.B.B.]

Sunflower diseases. Sunflowers are subject to destructive diseases throughout the world.

Causal organisms. Rust (*Puccinia helianthi*), caused by an air-borne fungus, reduces seed yields and may kill the plants. Root rot and wilt, caused by a soil-borne fungus, *Sclerotinia sclerotiorum*, affects sunflowers and also many other broad-leaved plant species. Other sunflower diseases caused by soil-borne fungi include downy mildew, *Plasmopara halstedii*, and leaf mottle, *Verticillium albo-atrum*, which are world-wide in distribution, and black root rot, *Sclerotium bataticola*, which is most common in tropical and subtropical regions. See FUNGI.

The virus diseases of sunflowers include mosaics and aster yellows (see PLANT VIRUS). They are still somewhat restricted in their geographical distribution.

Control measures. Chemical control of both rust and leaf mottle is possible, but it is not economically feasible. Cultural measures, such as weed removal and rotation with other crops which are d ease resistant, help reduce losses from most of sunflower diseases, and these measures are the controls available against some of them. Sunfl ers resistant to rust have been produced in Ma toba, Canada, and varieties resistant to downy m dew, leaf mottle, and aster yellows are being veloped there. [W.E.

Bibliography: See AGRICULTURAL SCIE (PLANT); PLANT DISEASE.

Sunlamp

A mercury-vapor gas-discharge tube used to p duce ultraviolet radiation for therapeutic or c metic purposes. The envelope of the tube is ma of quartz, since this is much more transparent ultraviolet radiation than glass. Ultraviolet rad tion from a sunlamp extends from visible vi light at about 3800 angstroms (A) to the tra mission limit of quartz at about 2200 A. Stro emission at about 2967 A is desirable for treati rickets. Longer wavelengths, approaching the vi ble spectrum, are more desirable for giving a g suntan in a reasonable time. The wavelength ultraviolet radiation can be controlled to a gr extent by the choice of material for the envelo of the tube, by using special glass filters, and using fluorescent sunlamps. In fluorescent su lamps the inside of the long glass tube is coat with a phosphor that gives radiation in the desir range when excited by ionization of the mercu vapor in the tube. See ULTRAVIOLET LAMP. [J.M

Sunshine-duration transmitter

A device for transmitting recordable informati concerning the number of hours the sun is visib each day compared with the number of hours possible sunshine. These instruments must be sen tive to direct sunshine, but not to sky brightne even though a cloudy sky near noon might brighter than direct sunlight near sunrise or su set. The Campbell-Stokes instrument uses a cle glass sphere to focus an image of the sun on paper card held in a bowl concentric to the gla sphere. The sun's image burns a trace in the pap card to provide a record. Near sunrise and suns the image may not be bright enough to burn th paper. The Maring-Marvin recorder is a clea bulb and black-bulb differential gas thermomete encased in a clear glass vacuum jacket to isolate from ambient air. In direct sunlight the warme black bulb moves a mercury column in a straig connecting tube between the bulbs toward contact embedded in the tube. The contact closure is use to record the presence of sunshine. Very bright sk near noon on an overcast day may cause this devic to indicate a visible sun in error. The above-men tioned defects are overcome in the photoelectri sunshine recorder adopted by the U.S. Weathe Bureau. In this, two self-generating photoelectri cells are connected in opposition to control a re lay. One of the cells is protected from direct sun light by a shade ring; the oth exposed to both

gime is being based on the slip velocity and mperature jump boundary conditions and on the avier Stokes equations.

Transition flow. No simple formulation for the ansition regime has yet been developed, except r the fundamental Maxwell-Boltzman equation self. Under some situations this equation can be olved for arbitrary values of λ/L, including those orresponding to transition flow conditions These pecial solutions, together with a large number of xperimental results of considerable generality, ndicate that simple interpolation between slip low on the one hand, and free molecule flow on the ther hand, will usually suffice for the transition low regime. See AERONAUTICAL ENGINEERING.

[S.A.S.]

... and M. Summerfield ...

lecular Flow of Gases, 1956.

Supercharger

An air pump or blower in the intake system of an internal combustion engine. Its purpose is to increase the air charge weight and power output ...

crease of altitude

Surface engines. For stationary, automobile, and marine duty, positive displacement blowers of the piston or Roots type are generally used, driven from the engine shaft Because the volume delivery of these types varies linearly with the engine speed, the cylinder pressure and shaft torque are reasonably constant throughout the speed range Typically, a gain in air charge and horsepower of perhaps one third may be realized over a naturally aspirated engine of the same size However, adiabatic compression of the intake air by the supercharger raises the air charge temperature which, in the spark ignition engine, promotes detonation and requires higher octane fuel Hence, except for racing cars, a supercharger is seldom used on spark ignition engines for mobile ground service

On diesel or compression ignition engine applications, the increased air content produced by the supercharger allows the engine to burn more fuel and produce greater horsepower without creating excessive pressures inside the cylinder. Also the supercharger enables a diesel engine to operate efficiently in the two-cycle mode. These two improvements have decreased the size and weight of the diesel engine so that it compares favorably with the spark ignition engine in these respects

Aircraft engines. To enable a piston type of internal combustion aircraft engine to develop its rated sea level horsepower at altitude, a supercharger must be used to increase the compression and weight of the intake air charge. Centrifugal compressors are used for this duty because of their relatively small size for a given capacity, and are driven either by a gear drive (from the crank shaft or by a gas turbine powered from the engine exhaust.

With the gear drive, a lower ratio is commonly ... and medium altitudes with a change to ... ed. During ... ry planes it was common practice to use a gear-driven second stage, supplemented at high altitude by a turbine-driven first stage, the exhaust feed to the turbine being bypassed at low altitudes. With two stages, it was usual to reduce the air charge temperature rise by use of an intercooler or an aftercooler (named according to whether the cooler was located between, or downstream of, the two stages).

A more recent special application is the turbo-compound engine, in which the turbine and compressor shaft have not only an exhaust turbine drive but also a gear connection with the main engine shaft, so that any excess energy from the turbine beyond that absorbed by the compressor can go into the shaft as additional drive horsepower. All these aircraft turbine installations have the advantage that the pressure ratio across the turbine tends to increase with altitude [F.C.MO]

Superconductivity

A property of certain metals, alloys, and chemical compounds at temperatures near absolute zero by virtue of which both their electrical resistivity and their magnetic permeability become immeasurably small

The transition of a metal from normal properties to superconducting properties depends principally upon (1) the temperature and (2) the magnetic field at the surface of the metal The superconductive state of the metal exists for temperatures less than the transition temperature T_c and for magnetic fields less than the critical field H_c, as shown in Fig. 1. The threshold curve dividing the normal from the superconductive states is roughly parabolic for all superconductors. To within a few per cent accuracy, $H_c = H_0 [1 - (T/T_c)^2]$, where H_0 is the critical field at absolute zero

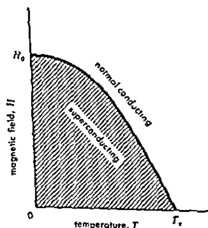


Fig 1 The superconductive state in the magnetic field-temperature plane

action parameters, also known as accommodation coefficients. These are

$$\alpha = \frac{e_i - e_r}{e_i - e_w} \qquad \sigma = \frac{\tau_i - \tau_r}{\tau_i} \qquad \sigma' = \frac{p_i - p_r}{p_i - p_w}$$

where p_i, τ_i, and e_i are the fluxes of normal momentum, tangential momentum, and energy incident on the surface; p_r, τ_r, and e_r are the fluxes of these quantities reemitted; while p_w, τ_w, and e_w are the fluxes that would be reemitted for the case of diffuse reflection. Measurements of these quantities for air in contact with typical engineering surfaces are generally in the range 0.8–1.0; however, very low values for α have also been reported, particularly for light gases in contact with very clean crystal surfaces. Utilizing the accommodation coefficient concept and carrying out the indicated calculations leads to basic formulae for net pressure p, shear stress τ, and heat flux q for a surface element at temperature T_w inclined at an angle of attack θ, to an incident free molecule flow of velocity V, temperature T, and density ρ. These are

$$p = \frac{\rho V^2}{2S^2}\left\{\left(\frac{2-\sigma'}{\sqrt{\pi}} S\sin\theta + \frac{\sigma'}{2}\sqrt{\frac{T_w}{T}}\right)e^{-S^2\sin^2\theta} + [(2-\sigma')(S^2\sin^2\theta + \tfrac{1}{2}) + \frac{\sigma'}{2}\sqrt{\frac{\pi T_w}{T}}\, S\sin\theta][1 + \operatorname{erf}(S\sin\theta)]\right\}$$

$$\tau = \frac{\sigma\rho V^2\cos\theta}{2\sqrt{\pi}S}\{e^{-S^2\sin^2\theta} + \sqrt{\pi}\, S\sin\theta\,[1 + \operatorname{erf}(S\sin\theta)]\}$$

$$q = \alpha\rho RT\sqrt{\frac{RT}{2\pi}}\left\{\left[S^2 + \frac{\gamma}{\gamma-1} - \frac{\gamma+1}{2(\gamma-1)}\frac{T_w}{T}\right] \times (e^{-S^2\sin^2\theta} + \sqrt{\pi}\, S\sin\theta[1 + \operatorname{erf}(S\sin\theta)]) - \tfrac{1}{2}e^{-S^2\sin^2\theta}\right\}$$

where $S = V/\sqrt{2RT}$ is known as the molecular speed ratio, R is the specific gas constant, and

$$\operatorname{erf}(S\sin\theta) = \frac{2}{\sqrt{\pi}}\int_0^{S\sin\theta} e^{-x^2}\,dx$$

is the error integral. By straightforward integration these results can be used to predict over-all aerodynamic and heat transfer characteristics for arbitrary convex configurations.

At somewhat higher densities the effect of molecular collisions begins to be manifest. Molecules [illegible]

be shown to be of order L/λ and is also sensitive to the Mach number and the temperature conditions.

Slip flow. This is the flow regime of only slight rarefaction, corresponding in the aerodynamic case to flight at altitudes of the order of 20–50 miles. Noncontinuum effects can be thought of in terms of small corrections to ordinary continuum flow. The characteristic dimension for many cases is the boundary layer thickness δ rather than the body dimension L. For low speed flow $\delta/L \sim 1/\sqrt{Re}$, so that the significant Knudsen number

$$K = \lambda/\delta \sim M/\sqrt{Re}$$

For high speed flow δ becomes dependent on the Mach number, the local heat transfer conditions, and the location on the aerodynamic body, so that a variety of different Knudsen numbers may become relevant.

The term slip flow arises from the phenomenon of slip, according to which a rarefied gas adjacent to a surface does not adhere rigidly to it but rather has a finite velocity, known as the slip velocity and determined by the local stress and temperature gradients. Physically, this phenomenon arises because the gas layer immediately adjacent to the surface is composed of molecules of which one-half are originating in the gas exterior to the surface and one-half are just being emitted from the surface. There is thus a discontinuity in the velocity and temperature. Mathematical analysis leads to a formulation for the velocity in the form

$$u_0 = \frac{2-\sigma}{\sigma}\lambda_0\left(\frac{\partial u}{\partial y}\right)_0 + \frac{3}{4}\frac{\nu_0}{T_0}\left(\frac{\partial T}{\partial x}\right)_0$$

where the subscript 0 refers to conditions at the surface, x is the coordinate along the surface in the direction of the flow, y is the coordinate normal to the surface, and u is the gas velocity in the x direction. The first term on the right side of this equation arises from the applied shear stress, while the second term, called thermal creep, arises from the temperature gradient in the direction of flow. The quantity $(2-\sigma)/\sigma$ is actually only a first approximation to a more complicated function of σ. The entire expression is applicable only for infinitesimal λ, $\partial u/\partial y$ and $\partial T/\partial x$. The corresponding thermal condition for the temperature jump is

$$T_0 - T_S = \frac{2-\alpha}{\alpha}\,\frac{2\gamma}{\gamma+1}\,\frac{\lambda_0}{Pr_0}\left(\frac{\partial T}{\partial y}\right)_0$$

where T_S is the actual surface temperature and Pr is the Prandtl number. This relation is subject to the same restrictions as the slip velocity condition. These relations express true noncontinuum effects which, by themselves, have the effect of reducing skin friction and heat transfer. Since they are of importance only at appreciable values of $M/\sqrt{Re}$ (or some other relevant Knudsen number), they are often obscured by continuum effects, for example, interaction effects arising from a combination of low Re and high M.

The situation with respect to the relevant flow equations is not quite so simple. Various modifications to the usual continuum Navier-Stokes equations have been proposed. However, in a series of experimental and theoretical investigations, none of these suggested modifications has so far turned out to be superior to the Navier-Stokes equations. Thus most present analytical w[illegible]k 'n the slip flow



Fig 3 Magnetic field distribution around a superconducting cylinder.

This depth is called the penetration depth λ. At absolute zero, $\lambda \equiv \lambda_0 \approx 5 \times 10^{-6}$ cm. The penetration depth varies with temperature approximately as $\lambda = \lambda_0[1 - (T/T_c)^4]^{-1/2}$ Colloidal particles, fine wires, and films having dimensions below 10^{-5} cm do not exhibit perfect diamagnetism even though they are still perfect conductors. Particles this small have critical fields H_c as much as 40 times higher than those in bulk materials

...

...ducting body depend on its geometry and orientation with respect to the applied field. Figure 3 illustrates the magnetic field distribution around a very long superconducting cylinder when placed in either of two orientations in a uniform field. The field, when along the cylinder as in Fig. 3*a*, is uniform at all points on the surface and is the same as the applied field for all values of that field When, however, the cylinder is placed perpendicular to the applied field, the field strength is doubled at the "equator" (position 2 in Fig 3*b*) and correspondingly decreased at the "poles" (position 1).

Now as the applied field is increased in the longitudinal case of Fig 3*a*, it is clear that the critical field H_c will be reached simultaneously at all points on the surface of the superconductor, as indicated in Fig 4 In this case, the entire specimen will become normal at once, and the sharply defined critical field curve of Fig 1 can be determined However, in the transverse orientation, the field at the equator reaches H_c for an applied field of $H_c/2$, while the field at the poles is still zero Further increase of applied field leaves the field at the equator const... enough (H_c ... out the cylir... netic transi... occurs, as revealed by detailed mapping of the fields at the surface At field strengths above $H_c/2$, magnetic flux starts to penetrate the cylinder via lamellae parallel to the applied field distributed more or less uniformly over the cross section. The field in each of these normal lamellae is just H_c. When the applied field is increased, the number of lamellae increases, but the field within each remains H_c. As the applied field reaches H_c, the lamellae increase to cover the entire cross section, and the cylinder becomes normal When the applied field is now reduced below H_c back to zero, this process is reversed The reversibility of this magnetizing process is an important distinction between perfect conductivity and superconductivity. If the cylinder were merely perfectly conducting, the magnetic flux, once having penetrated, would not again leave. For a discussion of a switching device based on this principle, *see* CRYOTRON.

Magnetization is not exactly reversible; there generally remains some flux threading through a superconducting specimen when an applied supercritical field is reduced to zero. The amount of flux "locked-in" increases with degree of chemical and physical imperfection, and may be as much as half the flux at the critical field for alloys or badly strained specimens.

Intermediate state. This transition mixture of normal and superconducting regions is termed the *intermediate state*. The detailed structure of the mixture depends on the energy required to create additional interfaces between normal and superconducting regions This interfacial or surface energy is akin to the surface tension between a liquid and its vapor, tending to minimize the area of the liquid-vapor boundary. The configuration of the intermediate state at any applied field represents a balance between the surface energy required to expand the superconducting regions and the magnetic energy gained by expelling the field from the normal regions

The surface energy plays a leading role in other aspects of the supercond...

... of one phase in another.

Most of the physical properties of metals have been studied on both sides of the superconducting transition to see if any of the differences observed would lead to better understanding of the nature of superconductivity Table 2 lists a number of such properties of pure tin above and below T_c in both the normal and superconducting states The properties in the normal state below T_c were measured in the presence of a magnetic field strong enough to destroy superconductivity The magnitude of

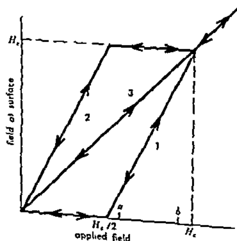


Fig. 4 Magnetic field at the surface of a cylinder placed in a uniform magnetic field The numbers 1, 2, 3 refer to the surface positions 1, 2, 3 in Fig 3.

Table 1. Transition temperatures and critical fields at absolute zero of the superconducting elements and some superconducting compounds

Substance	T_c, °K	H_0, oersteds
Al	1.20	106
Ti	0.39	20
V	5.03	1310
Zn	0.54	53
Ga	1.09	59
Zr	0.55	47
Nb	9.2	2000
Tc	11.2	300–400
Ru	0.47	46
Cd	0.54	29
In	3.40	278
Sn	3.72	305
La	5.9	1600
Hf	0.37	
Ta	4.39	780
Re	1.70	201
Os	0.71	65
Hg	4.16	413
Tl	2.39	171
Pb	7.22	807
Th	1.37	131
U	0.68	
Nb_3Sn	18.1	
V_3Si	17.1	
NbN	15–16	
U_6Fe	3.86	
CuS	1.6	
NaBi	2.2	
TlMg	2.75	220

Each superconductor is characterized by its pair of values (H_0, T_c). Table 1 lists these values (where known) for the 23 superconducting metallic elements and for an illustrative few of the scores of superconducting compounds. There are no inviolate rules for predicting whether or not a metal will be a superconductor. Most metallic elements have been tested down to 0.1°K, and many to 0.05°K or less. None of the alkali metals, noble metals, ferromagnetic metals, or antiferromagnetic metals are superconductors, although examples from each of these classes are found in superconducting compounds. Nearly all crystal classes are represented. In general, superconductivity is associated with high room-temperature resistivity. A further generalization is that superconductivity is most apt to occur in elements or compounds having 3, 5, or 7 valence electrons per atom.

EXPERIMENTAL ASPECTS

Zero resistivity. When the temperature of a superconductive metal is dropped below the transition temperature, the electrical resistance drops very abruptly. This sudden decrease of resistance was the basis of the discovery of superconductivity by K. Onnes in 1911 (Fig. 2). Within a small temperature interval (less than 0.001°K for some specimens near T_c), the resistivity decreases by a factor of at least 10^{11} to a value of less than 10^{-20} ohm-centimeters. A current induced in a closed loop (for instance, by withdrawal of a permanent magnet from within the loop) will decay in a fraction of a second if the loop is a normally conducting metal such as copper. If the loop is superconducting, however, the current will persist almost indefinitely. Persistent currents have been observed with no measurable decay for 18 months.

Electric currents are carried in superconductive circuits without resistive loss so long as the currents are not too large. A current flowing through a conductor establishes a magnetic field. The magnetic field H at the surface of a straight cylindrical wire of radius R carrying a current I is $H = I/5R$, where H is in oersteds, I in amperes, and R in centimeters. Normal conduction reappears when I is so large that $H > H_c$ of the wire. This is called the Silsbee effect, and it provides a practical limitation in the use of superconducting circuits. Nevertheless, the maximum currents which can be carried are sizable: a superconducting lead wire 2 mm in diameter can carry about 250 amp at 4°K.

Magnetic properties. Nothing in the previous remarks about the zero resistivity of superconductors anticipates their other basic property: zero permeability. W. Meissner in 1933 discovered that the magnetic field distribution near a superconductor changed as the superconductor went through its transition temperature (*see* MEISSNER EFFECT). Such change would not be expected from a perfect conductor, within which there can be no electromotive forces. Absence of these electromotive forces implies a constant interior magnetic field and, consequently, a fixed exterior field, which is contrary to observation in the case of a superconductor. The correct explanation lay in the complete absence of any internal magnetic induction. A normal material on becoming superconducting expels whatever magnetic flux it contains, or in other words becomes perfectly diamagnetic. *See* DIAMAGNETISM.

The statement that a magnetic field does not penetrate a metal in the pure superconducting state is not precisely true. Currents exist on the surface of the superconductor when an external field is present. These surface currents screen the interior by creating internal fields which exactly cancel the external field throughout the body of the superconductor. Obviously, these currents cannot flow on a two-dimensional surface, but require some depth in which to effect complete screening.

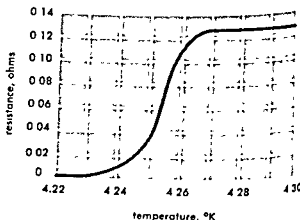
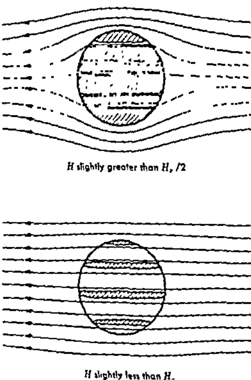


Fig. 2. Resistance of mercury as a function of temperature. (After K. Onnes, 19



H slightly greater than $H_c/2$

H slightly less than H_c

Fig 5. Magnetic field distribution around a cylinder in the intermediate state in a uniform transverse applied field.

Careful calorimetric measurements have shown that the portion of the jump due to the superconducting electrons varies with temperature as $e^{-bT_c/T}$, where *b* is a con...

0 6

hea

bei

the ...

... wave absorption experiments. Electrons can also be excited across an energy gap by receiving radiant energy from an electromagnetic field At low frequencies, a current in a superconductor flows with no energy loss whatsoever. At optical frequencies, however, currents lose energy at the same rate as in normal conductors. This loss may be regarded as arising *in part* from excitation of superconducting electrons to the normal state Absorption experiments in the infrared portion of the spectrum have shown that the frequency ν at which loss begins is given by $h\nu = 3.5kT_c$, where *h* is Planck's constant

Both calorimetric and high frequency data are consistent with a model having an energy gap $\epsilon \approx 3kT_c$.

Isotope effect. In 1950, C. A. Reynolds, B Serin, W H Wright, and L B Nesbitt discovered that the superconducting transition depends on the isotopic mass of the metal ions. E. Maxwell made the same discovery independently. The relations these men found, which have since been verified for a number of superconducting metals, were that for a given element the product $T_c\sqrt{M}$ is constant (M = isotopic mass of metal ion), and that H_0/T_c is the same for the various isotopes. Now the only effect of change in isotopic mass can be to change the vibrational pattern of the metal ions of the lattice. The fact that T_c depends on M was clear evidence that superconductivity arises somehow from an interaction between conduction electrons and lattice vibrations.

THEORY

The theories of superconductivity can be divided into two groups. The first group is largely phenomenological in nature, correlating a mass of experimental evidence into a coherent description from which indeed flows some insight into underlying mechanisms. The phenomenological theories make no pretense of explaining the origin of superconductivity on an atomistic basis The second group is more fundamental in inquiry and seeks to develop from elemental principles the basic mechanisms responsible for the observed phenomena. Progress in this second group has been slow, for a fundamental theory of superconductivity has been found to involve some of the most difficult problems in the quantum theory of solids.

Two-fluid model. The two-fluid model of superconductivity introduced by C. J. Gorter and H B G. Casimir in 1934 reproduces the principal thermodynamic properties of *superconductors.* This theory regards the conduction electrons in a *superconducting metal as* constituting two interpenetrating noninteracting electronic fluids. One, the superfluid component, does not interact with ...

... densed state of zero entropy, incapable of energy loss or transport of heat ...

other ...

cond...

ing-current resistance and heat conduction. The relative concentration of these fluids varies with temperature so that the normal fraction X is zero at $T = 0$ and unity at $T = T_c$, while the superconducting fraction $1 - X$ correspondingly varies from 1 to zero

The free energy difference ΔF between the normal and the superconducting phases at $T = 0$ is $VH_0^2/8\pi$, where V is the volume of the superconductor. At any other given temperature $\Delta F = \frac{1}{2}\gamma T^2(X^{\alpha} - 1) + (VH_0^2/8\pi)(1 - X)$. Here γT is the normal electronic specific heat, and α is a constant chosen to fit the experimental data. The quantity X must have that value X_e which makes ΔF a minimum if the two phases are to be in thermodynamic equilibrium. When $\alpha = \frac{1}{2}$, $X_e = (T/T_c)^4$, and $T_c = H_0/\sqrt{2\pi\gamma}$ Specific heat, entropy, and latent heat of transition may all be calculated from this two-fluid model with results which agree mod-

Table 2. Changes in various physical properties of pure* tin between normal (N) and superconducting (S) states (T_c = 3.73°K)

Property or effect	Above T_c	Below T_c			
	T = 3.8°K	T = 3.6°K		T = 2°K	
	N	N	S	N	S
Large change					
Resistivity, ohm-cm × 10^{10}	5	5	0	5	0
Permeability, gauss/oersted	1	1	0	1	0
Thermal conductivity, watts/(cm)(°K)	55	55	54	39	23
Specific heat, cal/(mole)(°K × 10^4)	60	60	73	13	13
Thermoelectric power, volts/°K × 10^9	18	15	0	2 5	0
Small change					
Volume V, $(V_S - V_N)/V_N$	0	0	5×10^{-9}	0	8×10^{-7}
Elastic modulus E, $(E_S - E_N)/E_N$	0	0	3×10^{-7}	0	3×10^{-6}

* The magnitudes of many of these physical properties depend on purity and crystallographic orientation and may even vary widely from one sample to another having the same nominal purity.

changes noted in the table are exemplary of other metals, although the magnitudes vary widely with purity and element. No changes were observed in optical and infrared reflectivity and absorptivity, x-ray and neutron scattering, beta-ray and positron absorption, and photoelectric and field emission.

Thermal properties. Since electrical conductivity is so extraordinary in superconductors, the question arises whether heat conductivity also is extraordinary. Heat is conducted through a metal partly by electrons roaming freely through the metal and partly by thermal vibrations of the ions forming the crystalline lattice. At low temperatures above 1°K, electronic heat conduction dominates. In general, the thermal conductivity of a pure metal rises with decreasing temperature, passes a maximum around 10°–20°K, and then decreases proportionately with temperature When a metal becomes superconducting, its thermal conductivity drops appreciably as might be expected, since the nature of the electronic conduction process also changes markedly. The drop in thermal conductivity suggests that some of the conduction electrons do not carry heat in the superconducting phase Below 1°K, the heat conduction in the normal phase may be hundreds of times larger than that in the superconducting phase Lead wires have been used employing this fact in a "heat switch" which is "open" when superconducting but may be closed by a supercritical magnetic field

The addition of only 10% of bismuth to lead greatly reduces the low-temperature thermal conductivity and also makes the superconducting state conduct better than the normal state (see Fig. 6). For additional details on thermal conductivity of metals, *see* CONDUCTION (HEAT).

There is a sudden jump in the specific heat of a superconductive metal when the metal becomes superconducting. The magnitude of this jump ΔC can be related thermodynamically to the slope $\partial H_c/\partial T$ of the critical field curve at T_c (Fig 1):

$$\Delta C = \frac{VT_c}{4\pi}\left(\frac{\partial H_c}{\partial T}\right)^2_{T_c}$$

where V is the volume of the superconductor. There is a latent heat Q evolved in transitions from the normal to the superconducting phase at temperatures below T_c:

$$Q = \frac{VT_cH_c}{4\pi}\frac{\partial H_c}{\partial T}$$

where H_c is the critical field at temperature T. Thus, the principal thermal properties of a superconductor can be determined directly from measurements of the critical field curve. That this is possible follows from the fact that the superconducting and normal phases are thermodynamically distinct and that reversible transitions can be made between them

Experiments aiding theory. Three types of experiment have been particularly important in the stimulus they have given to theories of superconductivity They are specific heat measurements, studies of ultra-high-frequency absorption, and the discovery of the isotope effect.

Specific heat measurements. The specific heat of a metal at low temperatures is in part due to the specific heat of the conduction electrons in the metal, and in part due to the thermal vibrations of the lattice (*see* LATTICE VIBRATIONS; SPECIFIC HEAT OF SOLIDS). The latter remain essentially unchanged through the superco • ing transition.

aximum, termed the Fermi energy, E_F. A fraction these electrons (of the order of kT/E_F) are thermally excited to higher energies than E_F at temperatures higher than 0°K. The thermal excitation of these electrons gives rise to a normal electronic specific heat proportional to T.

In the superconducting state, however, the electronic specific heat is much smaller and varies exponentially with temperature as described earlier. This suggests that fewer electrons are thermally excited when $T < T_c$. This in turn suggests that a definite energy of the order of kT_c is required to excite an electron near $T = 0$; that is, there are no (or few) states of possible energy in a gap of width $E \sim kT_c$ near E_F. Thermal excitation of electrons across this energy gap at temperatures $0 < T < T_c$ leads to an electronic specific heat and thermal conductivity, both varying exponentially with temperature.

It can be shown that such an energy-gap model leads directly to the Meissner effect, thus accounting not only for thermal effects but for magnetic properties of superconductivity as well. The energy-gap model therefore provides a realistic picture of the energy spectrum of electrons in a superconductor. One of the tasks of a fundamental theory of superconductivity must be to explain the existence of an energy gap in this spectrum.

Quantum theories. A complete theory of superconductivity would be expected to start with the basic equations of the quantum theory and end by explaining all the observed features of superconductivity. Development of such a theory will be enormously difficult, although important steps toward it have been taken. The root of the difficulty lies in the small energy difference between the normal and the superconducting states, amounting to only 10^{-8} electron volts (ev) per conduction electron. This difference is to be compared with the energies of electrons responsible for normal conduction, of the order of 10 ev per electron. No theory that has yet been devised permits such a calculation of electronic energies to one part in 10^9. Instead, the most successful theoretical approaches have pretended that the energies and possible energies of the normal conduction electrons in a metal are known precisely and have then inquired what small changes in these energies would give rise to superconductivity. These inquiries have been guided by the phenomenology of superconductivity, in particular by the realization that a gap in possible electron energies would account for the thermal and conduction properties and the Meissner effect.

The remaining remarks here will deal with the theory first published in 1957 by J. Bardeen, L. N. Cooper, and J. R. Schrieffer. This theory has been most successful in describing such diverse behaviors of superconductivity as temperature dependence of specific heat, of penetration depth, and of ultrasonic absorption.

The Bardeen-Cooper-Schrieffer theory received its initial impetus from the discovery of the isotope effect described earlier. The mass dependence of the superconducting transition suggested strongly that superconductivity is associated with thermal vibrations of the lattice ions. The long-range order of the superconducting electrons is evidence of some form of long-range interaction between them. The Bardeen-Cooper-Schrieffer theory shows how lattice vibrations are involved. According to this theory, the interaction arises in the following way: An electron moving through a lattice collides with it and subtly changes the vibrational pattern of the lattice. This small change is in turn communicated to another electron traveling in a direction opposite to the first. In this fashion, which can be properly described only in quantum mechanical language, electrons tend to interact in pairs and in such a way that their energy is reduced. When this interaction is summed over all pairs of electrons traveling all directions in the metal, the net result is an over-all lowering of electron energies sufficient to leave the energy gap sought by the theory. The principal phenomena of superconductivity then can be deduced from the magnitude and temperature dependence of this energy gap.

Theoreticians are in general agreement that the Bardeen-Cooper-Schrieffer theory is basically correct, even though incomplete in certain basic aspects, in particular in describing the nature of the electron-lattice interaction. The theory clearly represents a major step toward a complete understanding of the nature of superconductivity.

[M.D.F.]

Bibliography. S. Fluegge (ed.), *Handbuch der Physik*, vol. 15, 1956; C. J. Gorter (ed.), *Progress in Low Temperature Physics*, vol. 1, 1955; M. T. F. von Laue, *Theory of Superconductivity*, 1952; F. London, *Superfluids*, vol. 1, 1950; D. Shoenberg, *Superconductivity*, 1952.

Superfluidity

The strange frictionless flow of matter observed in liquid helium below 2.186°K, the λ point of helium; also, the flow without resistance of electric current at sufficiently low temperatures in certain solids (superconductivity). Liquid helium can flow without any apparent friction through slits so small that ordinary liquids and even gases can hardly pass through. This strange fluid state also exhibits a very high and unusual heat conductivity, a unique kind of film flow, a thermal wave motion called second sound, and a thermomechanical effect resulting in certain cases in the so-called fountain effect. In a superconductor an electric charge can pass without any measurable voltage between its ends. The extent to which the flow of electric current in a superconductor is frictionless has been demonstrated by an experiment in which a persistent electric current has been observed to flow without measurable decrease in amplitude around a superconducting lead ring for over 2 years. The superfluid state is of particular interest to physicists because it represents a long range order in the motion of the particles (order in momentum ...

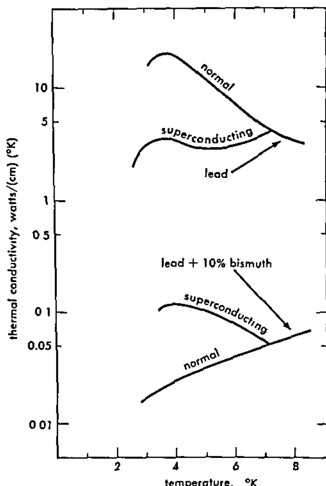


Fig. 6 The thermal conductivity of pure and alloyed lead in the normal and superconducting states, as a function of temperature.

erately well with experimental data This is not too surprising, since some of the physical behavior of superconductors was built into the formulation of the model (as in the choice of α).

London equations. The phenomenological theory developed by F. London and H. London in 1935 describes the electromagnetic behavior of superconductors. The basic equations of the theory are essentially postulates, justified by experimental confirmation of the behaviors they imply. The London equations assert that the supercurrent density at a point is always determined by the magnetic field at that point, whereas in normal conduction the current is determined by the local electric field. The equations are

$$-\text{curl}\left(\frac{4\pi\lambda^2}{c}J_s\right) = H \qquad (1)$$

and

$$\frac{\partial}{\partial t}\left(\frac{4\pi\lambda^2}{c^2}J_s\right) = E \qquad (2)$$

Here J_s is the density of the supercurrent, H and E are the respective magnetic and electric fields, c is the velocity of light, and λ is the penetration depth discussed earlier. When the two-fluid assumption is made that the total current density J is the sum of J_s plus the normal current density J_n and that the normal electrons obey Ohm's law ($J_n = \sigma E$, where σ = electrical conductivity), then the London equations for the total current become

$$-\text{curl}\left(\frac{4\pi\lambda^2}{c}J\right) = H + \frac{4\pi\sigma\lambda^2}{c^2}\frac{\partial H}{\partial t} \qquad (3)$$

and

$$\frac{\partial}{\partial t}\left(\frac{4\pi\lambda^2}{c^2}J\right) = E + \frac{4\pi\sigma\lambda^2}{c^2}\frac{\partial E}{\partial t} \qquad (4)$$

When (3) and (4) are combined with Maxwell's equations of electromagnetism, it is found that both H and E decrease exponentially from the surface into the interior. One of the striking successes of the London theory was the quantitative prediction it made about the penetration depth of these fields, namely that $\lambda = (mc^2/4\pi ne^2)^{1/2}$, where n is the effective number of free electrons per unit volume in the metal, m is the electronic mass, and e the electronic charge. Insertion of the appropriate numerical values led to $\lambda_0 \approx 2 \times 10^{-6}$ cm, a figure later substantiated experimentally. The magnetic and electric fields then are excluded from all but a very small fraction of a superconductor. Equations (3) and (4), therefore, lead to the Meissner effect. Equation (2) amounts to a statement of the perfect conductivity of superconductors: the electric field within is zero for static conditions. Thus, the London equations provide for the zero permeability and zero resistivity required for any theory of superconductivity.

The London theory has been useful in correlating the results of experiments, especially those dealing with very thin films and with high frequency effects. It predicted that a very small superconductor should have a much higher critical field than a large one; this was later confirmed experimentally The theory also correctly predicted that the electrodynamic behavior of a superconductor should be the same as that of a normal conductor at frequencies above about 10^{11} cycles per second It had been known for some time that superconductivity did not affect the optical properties of metals, but the frequency range in which normal conduction gave way to superconduction was not known

While the London theory has been a useful guide and generally leads to correct order-of-magnitude predictions, it has left something to be desired in quantitative prediction It further fails to provide for a magnetic field dependence of the penetration depth, observed by A. B Pippard This observation led Pippard to propose a generalization of the London equations based on a nonlocal relation between current density and magnetic field. Fundamental to Pippard's theory is the concept of coherence, the idea that the superconducting electrons exhibit long-range order over spatial regions extending 10^{-4} cm. This is to say that distances of this magnitude (which are very much larger than the penetration depth) are required before substantial changes can occur in the fraction of superconducting electrons

In a normal metal at $T = 0$°K, conduction electrons may have energies ra[illegible] from zero up to a

maximum, termed the Fermi energy, E_F. A fraction of these electrons (of the order of kT/E_F) are thermally excited to higher energies than E_F at temperatures higher than 0°K. The thermal excitation of these electrons gives rise to a normal electronic specific heat proportional to T.

In the superconducting state, however, the electronic specific heat is much smaller and varies exponentially with temperature as described earlier. This suggests that fewer electrons are thermally excited when $T < T_c$. This in turn suggests that a definite energy of the order of kT_c is required to excite an electron near $T = 0$; that is, there are no (or few) states of possible energy in a gap of width $E \sim kT_c$ near E_F. Thermal excitation of electrons across this energy gap at temperatures $0 < T < T_c$ leads to an electronic specific heat and thermal conductivity, both varying exponentially with temperature.

It can be shown that such an energy-gap model leads directly to the Meissner effect, thus accounting not only for thermal effects but for magnetic properties of superconductivity as well. The energy gap model therefore provides a realistic picture of the energy spectrum of electrons in a superconductor. One of the tasks of a fundamental theory of superconductivity must be to explain the existence of an energy gap in this spectrum.

Quantum theories. A complete theory of superconductivity would be expected to start with the basic equations of the quantum theory and end by explaining all the observed features of superconductivity. Development of such a theory will be enormously difficult, although important steps toward it have been taken. The root of the difficulty lies in the small energy difference between the normal and the superconducting states, amounting to only 10^{-8} electron volts (ev) per conduction electron. This difference is to be compared with the energies of electrons responsible for normal conduction, of the order of 10 ev per electron. No theory that has yet been devised permits such a calculation of electronic energies to one part in 10^9. Instead, the most successful theoretical approaches have pretended that the energies and possible energies of the normal conduction electrons in a metal are known precisely and have then inquired what small changes in these energies would give rise to superconductivity. These inquiries have been guided by the phenomenology of superconductivity, in particular by the realization that a gap in possible electron energies would account for the thermal and conduction properties and the Meissner effect.

The remaining remarks here will deal with the theory first published in 1957 by J. Bardeen, L. N. Cooper, and J. R. Schrieffer. This theory has been most successful in describing such diverse behaviors of superconductivity as temperature dependence of specific heat, of penetration depth, and of ultrasonic absorption.

The Bardeen-Cooper-Schrieffer theory received its initial impetus from the discovery of the isotope effect described earlier. The mass dependence of the superconducting transition suggested strongly that superconductivity is associated with thermal vibrations of the lattice ions. The long-range order of the superconducting electrons is evidence of some form of long-range interaction between them. The Bardeen-Cooper-Schrieffer theory shows how lattice vibrations are involved. According to this theory, the interaction arises in the following way: An electron moving through a lattice collides with it and subtly changes the vibrational pattern of the lattice. This small change is in turn communicated to another electron traveling in a direction opposite to the first. In this fashion, which can be properly described only in quantum mechanical language, electrons tend to interact in pairs and in such a way that their energy is reduced. When this interaction is summed over all pairs of electrons traveling all directions in the metal, the net result is an over-all lowering of electron energies sufficient to leave the energy gap sought by the theory. The principal phenomena of superconductivity then can be deduced from the magnitude and temperature dependence of this energy gap.

Theoreticians are in general agreement that the Bardeen-Cooper-Schrieffer theory is basically correct, even though incomplete in certain basic aspects, in particular in describing the nature of the electron-lattice interaction. The theory clearly represents a major step toward a complete understanding of the nature of superconductivity.

[M.D.F.]

Bibliography. S. Fluegge (ed.), *Handbuch der Physik*, vol. 15, 1956; C. J. Gorter (ed.), *Progress in Low Temperature Physics*, vol. 1, 1955; M. T. F. von Laue, *Theory of Superconductivity*, 1952; F. London, *Superfluids*, vol. 1, 1950; D. Shoenberg, *Superconductivity*, 1952.

Superfluidity

The strange frictionless flow of matter observed in liquid helium below 2.186°K, the λ point of helium; also, the flow without resistance of electric current at sufficiently low temperatures in certain solids (superconductivity). Liquid helium can flow without any apparent friction through slits so small that ordinary liquids and even gases can hardly pass through. This strange fluid state also exhibits a very high and unusual heat conductivity, a unique kind of film flow, a thermal wave motion called second sound, and a thermomech[illegible] ing ir[illegible] In a [illegible] witho[illegible] measurable voltage between its ends. The extent to which the flow of electric current in a superconductor is frictionless has been demonstrated by an experiment in which a persistent electric current has been observed to flow without measurable decrease in amplitude around a superconducting lead ring for over 2 years. The superfluid state is of particular interest to physicists because it represents a long-range order in the motion of the particles (order in momentum space)

instead of the usual order in the arrangement of the particles (order in position space) found near the absolute zero of temperature in all other substances. The superfluid state can be explained only in terms of quantum mechanics and represents a macroscopic or large-scale quantum state. In the case of liquid helium, this has been dramatically demonstrated by the theoretical prediction and the experimental observation of quantized rotational flow. Superfluidity in liquid helium is thought to be related to the condensation occurring at low temperatures in a theoretical, ideal, Bose-Einstein gas. Indeed, superfluidity is not observed in the rare isotope He^3 which, unlike He^4, obeys the Fermi statistics instead of the Bose-Einstein statistics. *See* BOSE-EINSTEIN STATISTICS; HELIUM, LIQUID; SUPERCONDUCTIVITY. [W.M.FA.]

Supergiant star

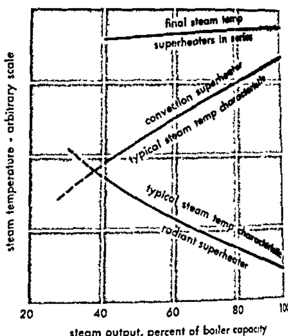
A member of the family containing the intrinsically brightest stars, populating the top of the Hertzsprung-Russell diagram (*see* STAR). Supergiant stars occur at all temperatures, from 30,000 to 3,000°K, and have luminosities ranging from 10^4 to 10^6 times that of the Sun. The hot supergiants have radii 20 times the Sun The cool supergiants are the largest known stars, reaching several thousand solar radii. Among the bright supergiants are Deneb (α Cygni), Rigel (β Orionis) and Betelgeuse (α Orionis). Because few have masses exceeding 20–40 Suns, the ratio of luminosity to mass is high. Thermonuclear energy sources will therefore be rapidly exhausted, hydrogen being consumed in a time scale of only 1 to 10,000,000 years The supergiants are thus young but rapidly evolving massive stars Nearly all are slightly variable in light and radial velocity, have highly turbulent and extended atmospheres, and show broadened spectral lines. [J.L GR]

Superheater

A component of a steam generating unit in which high-pressure steam, after it has left the boiler drum, is heated above the saturation temperature in order to increase the enthalpy of the steam before it is delivered to an engine or turbine. The amount of superheat acquired is anticipated by design, and depends primarily upon the extent of superheater surface installed, as well as on its location in the unit. The superheater may consist of several stages, the final temperature being attained in sections exposed to high-temperature portions of the gas stream, or to the combustion zone of the furnace.

The primary or initial stage of a superheater usually consists of a tube bank, formed of parallel loops set on relatively close spacing, which are swept by gases that have given up some heat to the final stage of the superheater. In such a bank of tubes the transfer of heat from the gases is predominantly by convection and the section is referred to as a convection superheater.

For the highest outlet steam temperature, the final stage of the superheater is located closer to the furnace. In some designs the tubes form part of the furnace enclosure, and in others project into the furnace cavity as loops or platens on wide lateral spacing. Heat transfer to the elements occurs primarily by radiation, and such arrangements are known as radiant superheaters.



When convection and radiant superheaters operate in series, final steam temperature is uniform over a wide range of rates of steam generation.

Since the temperature characteristics of these extreme types differ, as operating rate is varied on the unit, it is frequently advantageous to combine them in series for obtaining more uniform outlet steam temperature over a wide load range (as illustrated). Intermediate characteristics are approached when fuel conditions permit arranging the superheater bank for both effective radiation and convection transfer. *See* STEAM GENERATING UNIT; STEAM TEMPERATURE CONTROL. [F C E]

Superheterodyne receiver

A receiver that uses the heterodyne principle to convert the incoming modulated radio-frequency signal to a predetermined lower carrier frequency, called the intermediate-frequency (i-f) value. This is done by using a local oscillator that is tuned simultaneously with the input stage of the receiver, so that the oscillator frequency will always differ from that of the incoming carrier by the i-f value (*see* HETERODYNE PRINCIPLE)

With a fixed and favorably chosen i-f value, the i-f amplifier can efficiently provide the major portion of the amplification and selectivity required by the receiver. After amplification, the i-f signal is demodulated by the second detector to obtain the desired audio output signal A similar circuit arrangement is used in television and radar receivers to obtain the desired video output signal. For radio broadcast receivers the i-f value is usually 455 kilocycles *See* RADIO RECEIVER [J.M.R]

Supernova

A star that suddenly bursts into great brilliance, as a result of the greatest explosion known to mankind. A supernova probably forms when a whole star blows up instead of just blowing off a small amount of its atmosphere, as does a galactic nova (*see* NOVA). At maximum, the supernova is often as bright or brighter than the total light of the galaxy in which it occurs.

More than 50 supernovae have been observed, but only 3 of these are definitely known to have been members of our galaxy. Most of the others are located in external systems; there is doubt about a few of the older and less authenticated records. The faint rings of filamentary nebulae photographed with large telescopes may be the remains of supernovae, which appeared in our galaxy many centuries ago. It has been ...

... from galactic novae by their greater amplitudes, their bright absolute magnitudes, and their peculiar spectra.

Little information has been obtained about supernovae, they seem to fall into two general types. Type I is distinct ...

... ae instead.

Type I. A type I supernova has an absolute magnitude at maximum of at least −16. All members of the group have similar light curves, so uniform that it is possible to extrapolate the whole curve from fragmentary data. The curve shows a rapid rise followed by a rapid decrease for about a month, after which it begins to approach a slower but steady decline. The spectra are formed of extremely broad, overlapping bright lines, conspicuous just before maximum. Their widths suggest velocities of expansion of 10,000 km/sec. These lines have not as yet been satisfactorily identified. After maximum, narrower emission lines of neutral oxygen appear, which indicate a velocity of about 1000 km/sec. It is generally agreed that type-I supernovae are ...

... a slower decline than the supernovae of type I, and have a marked period of constant brightness on the decreasing slope of the curve. Before and at maximum, the spectra of type-II supernovae are continuous and show no emission, but about a month later, wide emission lines develop. They have been identified as due to hydrogen and nitrogen (N III).

Supernova of 1054. The three supernovae known to be members of our galaxy appeared in the years 1054, 1572, and 1604. The supernova of 1054 (CM Tauri) was observed by Chinese and Japanese astronomers and for 2 years was visible to the unaided eye. Surprisingly, no European records of it have been found. The nebulous remnants of the supernova began expanding at the time of the explosion and now, 900 years later, are seen as a spectacular nebula (see CRAB NEBULA).

The Crab nebula is still expanding at the rate of 0.21 seconds of arc per year. N. U. Mayall has determined from radial velocity measurements that the rate of expansion is 1300 km/sec. The Crab nebula coincides with a powerful continuous radio source, which probably is from the combination of the filaments and the central star, which is about sixteenth magnitude. J. H. Oort concludes that if the Crab nebula continues to expand at the same rate, about 30,000 years from now it will resemble the great Veil nebula in Cygnus. The speculation ...

... probably reached apparent magnitude −6, which at its distance of 4000 light years corresponds to an absolute magnitude of −16. If the star of the Veil nebula reached the same luminosity, its apparent magnitude must have been −9 and was brighter than the Moon at first quarter.

The supernova of 1572, β Cassiopeiae, known as Tycho's nova, reached a maximum of −4.0. No faint star has been identified in its position, and its range must have been more than 22 magnitudes. In 1956, R. Minkowski found faint nebulosity near the position and a radio source was observed in 1952 by Hanbury Brown and C. Hazard. β Cassiopeiae was almost certainly a type-I supernova, from the descriptions of its color and light curve.

The supernova of 1604, V843 Ophiuchi, known as Kepler's nova, has a well observed, type-I light curve, with a maximum at −2.2 magnitude. No faint star or radio source has been found in its position. However, a faint nebulous patch is seen nearby, and Minkowski obtained spectra of it in 1943. The spectrum is similar to that of the Crab nebula and other type-I supernovae. It shows emission lines of oxygen and nitrogen but is hydrogen-poor.

Another possible remnant of a galactic supernova of the past is the great hydrogen alpha ring of nebulosity that almost surrounds the constellation of Orion. Ernst Öpic called attention to it in 1953 and compared it with a similar nebula in the Large Magellanic Cloud, which he said is almost certainly the remains of an old supernova explosion. Other galactic supernovae may be found among the many bright temporary stars noted in early Oriental records. Unexplained radio sources also may be from supernova remnants. See VARIABLE STAR.

[M.W.M.]

Superposition, principle of

In classical wave theories (optics, acoustics), as in all theories characterized by linear homogeneous differential equations, the sum of any number of solutions to the equations is another solution. Thus (assuming one-dimensional waves for simplicity),

the amplitude of the resultant wave at a point x and time t is the linear superposition of the amplitudes of all waves reaching x at time t. This fact, known as the principle of superposition, often also is taken to mean that if each of $\psi_1(x)$ and $\psi_2(x)$ are possible waveforms at $t = 0$, then any linear combination $\psi(x) = c_1\psi_1 + c_2\psi_2$ is a possible waveform at $t = 0$. This latter version is not a consequence of the linearity of the differential equations, but rather is an affirmation of the belief that the waveform can be chosen arbitrarily at any initial instant. Conversely, if $u_n(x,t)$ is the solution which at $t = 0$ equals $u_n(x) \equiv u_n(x,0)$, and if $u_n(x)$ are a complete set of functions, then any wave $\psi(x,t)$ which at $t = 0$ equals $\psi(x)$ can be written in the form

$$\psi(x,t) = \sum_n c_n u_n(x,t)$$

where the constants c_n are determined from the equation at $t = 0$. *See* WAVE MOTION.

The preceding paragraph holds equally well for quantum theory, where the wave function $\psi(x,t)$ obeys the linear homogeneous Schrodinger equation In quantum theory, however, because the wave function represents physical states, the principle of superposition has a profound significance. In particular, if the observable A is certain to have value α_1 in the state represented by u_1, and is certain to have value α_2 in the state represented by u_2, then $c_1u_1 + c_2u_2$ represents a state in which measurement of the observable A is certain to yield either precisely α_1, or else precisely α_2. Thus, because probabilities are proportional to $|\psi|^2$, superposition accounts for phenomena which are difficult to understand from a classical viewpoint For example, suppose u_1 represents a beam which passes through only one of a pair of slits, and u_2 represents a beam emitted from the same source, but directed at the second slit. Then $\psi = c_1u_1 + c_2u_2$ is the wave function in a double slit experiment, with $|c_1|^2$, $|c_2|^2$ the relative intensities of the two beams, that is, $|c_1|^2$, $|c_2|^2$ are the probabilities that particles (photons in light beams, electrons in electron beams) reach the viewing screen via slits 1 and 2, respectively. The beam intensities can be made so low that only one particle reaches the screen each second, which particle (according to the classical viewpoint) must have come either through slit 1 or slit 2. Nonetheless, the probability of observing a particle at a point y on the viewing screen is $|\psi(y)|^2 = |c_1u_1(y) + c_2u_2(y)|^2$, wherein the two beams interfere. *See* QUANTUM MECHANICS. [E.C.]

Superposition theorem (electric networks)

This theorem may be stated as follows: in any linear bilateral network containing generators, the current flowing in any branch is the sum of the currents which would result from each generator acting independently, other generators being replaced at the time by their internal impedances. Linear bilateral networks contain elements that have linear current-voltage relationship and that transmit energy equally in either direction. Resistors, capacitors, and inductances are linear and bilateral. Electron tubes are nonlinear and unilateral Superposition is not permissible if the elements are not constant. For general theory of circuits *see* ALTERNATING-CURRENT CIRCUIT THEORY; DIRECT-CURRENT CIRCUIT THEORY.

The principle of superposition is one of the most important theorems in network analysis It is particularly useful in proving other theorems, but it is not often used for circuit calculations

When a linear bilateral network contains only generators of the same frequency, the actual current in any branch is the phasor sum (or the algebraic sum in the case of instantaneous values) of the individual currents due to the individual generators

When a linear bilateral network contains generators of different frequencies, the principle of superposition permits a solution to be obtained for each frequency separately. The individual currents due to the individual generators should be expressed in time functions (instantaneous expressions) before they are combined.

Figure 1 shows a network with two voltage generators of the same frequency. The internal impedance of generator E_1 is R_1 and that of generator E_2 is R_2 With S_1 and S_2 open, it is desired to find I_3, I_4, and I_5.

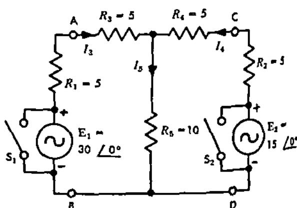


Fig 1 Network containing two voltage generators

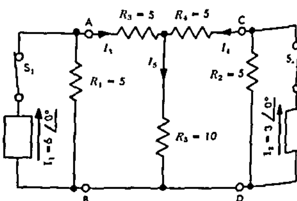


Fig 2. Network containing t generators

With generator E_2 replaced by its internal impedance R_2 (E_2 is made inactive by having S_2 closed), the currents due to E_1 are

$$I'_3 = \frac{30/0^\circ}{15/0^\circ} = 2/0^\circ \qquad I'_4 = -1/0^\circ \qquad I'_5 = 1/0^\circ$$

With generator E_1 replaced by its internal impedance R_1 (E_1 is made inactive by having S_1 closed), the currents due to E_2 are

$$I''_4 = \frac{15/0^\circ}{15/0^\circ} = 1/0^\circ \quad I''_3 = -0.5/0^\circ \quad I''_5 = 0.5/0^\circ$$

The actual currents due to both generators are

$$I_3 = I'_3 + I''_3 = 2/0^\circ - 0.5/0^\circ = 1.5/0^\circ$$
$$I_4 = I'_4 + I''_4 = -1/0^\circ + 1/0^\circ = 0$$
$$I_5 = I'_5 + I''_5 = 1/0^\circ + 0.5/0^\circ = 1.5/0^\circ$$

The power dissipated in R_5 only is

$$P_5 = I_5^2 R_5 = (1.5)^2 \times 10 = 22.5 \text{ watts}$$

It is not

$$I'_5{}^2 R_5 + I''_5{}^2 R_5 = 1^2 \times 10 + (0.5)^2 \times 10 = 12.5 \text{ watts}$$

That is, the principle of superposition does not hold for power, because power is a quadratic function instead of a linear function of current.

Figure 2 shows a network with two current generators of the same frequency. The internal impedance (shunt impedance) of generator I_1 is R_1 and that of generator I_2 is R_2. With S_1 and S_2 closed, it is desired to find I_3, I_4, and I_5.

With generator I_2 replaced by its internal impedance R_2 (I_2 is made inactive by having S_2 open), the currents due to I_1 are

$$I'_3 = 2/0^\circ \qquad I'_4 = -1/0^\circ \qquad I'_5 = 1/0^\circ$$

With generator I_1 replaced by its internal impedance R_1 (I_1 is made inactive by having S_1 open), the currents due to I_2 are

$$I''_4 = 1/0^\circ \qquad I''_3 = -0.5/0^\circ \qquad I''_5 = 0.5/0^\circ$$

The actual currents due to both I_1 and I_2 are

$$I_3 = I'_3 + I''_3 = 2/0^\circ - 0.5/0^\circ = 1.5/0^\circ$$
$$I_4 = I'_4 + I''_4 = -1/0^\circ + 1/0^\circ = 0$$
$$I_5 = I'_5 + I''_5 = 1/0^\circ + 0.5/0^\circ = 1.5/0^\circ$$

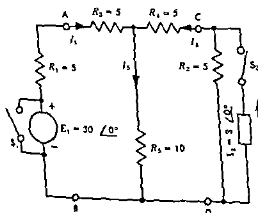


Fig 3 Network containing one voltage generator and one current generator

Figure 3 shows a network with two generators of the same frequency. The internal impedance (series impedance) of generator E_1 is R_1 and the internal impedance (shunt impedance) of generator I_2 is R_2. With S_1 open and S_2 closed, find I_3, I_4, and I_5.

With generator I_2 replaced by its internal impedance R_2 (S_2 open), the currents due to E_1 are

$$I'_3 = \frac{30/0^\circ}{15/0^\circ} = 2/0^\circ \qquad I'_4 = -1/0^\circ \qquad I'_5 = 1/0^\circ$$

With generator E_1 replaced by its internal impedance R_1 (S_1 closed), the currents due to I_2 are

$$I''_4 = 1/0^\circ \qquad I''_3 = -0.5/0^\circ \qquad I''_5 = 0.5/0^\circ$$

The actual currents due to both E_1 and I_2 are

$$I_3 = I'_3 + I''_3 = 1.5/0^\circ \qquad I_4 = I'_4 + I''_4 = 0$$
$$I_5 = I'_5 + I''_5 = 1.5/0^\circ$$

See NETWORK THEORY, ELECTRICAL [K.Y.T.]

Supersaturation

A solution is at the saturation point when dissolved solute in it crystallizes from it at the same rate at which it dissolves. Under prescribed experimental conditions of temperature and pressure, a solution can contain at saturation only one fixed amount of dissolved solute. However, it is possible to prepare relatively stable solutions which contain a quantity of a dissolved solute greater than that of the saturation value provided *solute phase* is absent. Such solutions are said to be *supersaturated*. They can be prepared by changing the experimental conditions of a system so that greater solubility is obtained, perhaps by heating the solution, and then carefully *returning* the system to or near its original state. The addition of solute phase will immediately relieve supersaturation. Solutions in which there is no spontaneous formation of solute phase for extended periods of time are said to be *metastable*. There is no sharp line of demarcation between an unstable and a metastable solution. In fact, the latter is poorly defined and much influenced by many factors such as mechanical shock and the presence of minute quantities of foreign materials. The process whereby initial aggregates within a supersaturated solution develop spontaneously into particles of new stable phase is known

...metric analysis because of the formation of many small crystals which tend to coprecipitate exces-

sive amounts of foreign ions by virtue of their great surface area. *See* EQUILIBRIUM, PHASE; GRAVIMETRIC ANALYSIS; *NUCLEATION*; *PRECIPITATION (CHEMISTRY)*; SATURATION OF SOLUTIONS. [L.C.]

Supersonic diffuser

A passive compressor in which gas enters at a velocity greater than the speed of sound, is decelerated in a contracting section, and reaches sonic speed at a throat.

Operating principle. Ideally, supersonic diffusion consists of an isentropic or constant total pressure compression or deceleration of the flow from the free-stream Mach number to a Mach number of unity (or sonic velocity) in the throat. Supersonic deceleration is accomplished by reducing the cross section of the streamtube. *See* DIFFUSER.

Practically, the achievement of isentropic compression is modified to varying degrees by viscous effects, shock-boundary-layer interactions, inlet starting limitations, or compression limits based on theoretical shock structures Such restraints and means for avoiding or minimizing their effects are the main problems in the design of efficient supersonic diffusers.

The transition from supersonic to subsonic flow occurs practically always through a normal shock. The associated losses, however, increase rapidly *with increasing Mach number. At speeds above* Mach 1.5, such losses become intolerable and a more refined compression system is required. Such a system consists of a gradual deceleration of the flow through a system of one or more weak oblique shock waves. Theoretically, isentropic deceleration can be achieved through an infinite number of weak waves

Basic types. Supersonic compression systems *can be categorized in three basic types (Fig. 1).* External-compression inlets have the supersonic diffusion taking place at or ahead of the cowl lip (or throat station) and generally employ one or more oblique waves ahead of the normal shock. Theoretically, pressure recovery increases with the *number of oblique shocks (Fig. 2).* For all external-compression inlets, there exists a compression limit based on an analysis of theoretical shock structures. Further, with increased compression, there is a concomitant increase in the flow turning angle, which is reflected in increasingly steeper cowl angles and thus cowl drag

Internal-compression inlets accomplish supersonic diffusion internally downstream of the cowl lip. Deceleration of the flow is produced by a number of weak reflecting waves in a gradually convergent channel. Surface angles are necessarily kept small to maintain small adverse pressure gradients on the boundary layer. This means a long diffuser. The most severe requirement on this type of inlet is dictated by the starting problem To start, the diffuser must swallow the shock and establish supersonic flow past the throat. To attain high performance levels and to attain operating contraction ratios in excess of a theoretical maximum starting value, variable geometry or vent perforations in the convergent duct are necessary, although they add complexity. With this type of inlet, there appears to be no limit on theoretical recovery and zero or low drag cowls are usually possible.

The third system is a combination of external and internal compression and appears to represent an effective compromise. By incorporating external compression, the variable-geometry requirements are considerably less than with the all internal system, and, through the use of internal contraction, the cowl drag is less and the potential total pressure recovery higher than with an all-external compression system.

Practical considerations. *Viscous effects* are an important factor in supersonic inlet performance The wall boundary layers must negotiate the high adverse pressure gradients that arise from abrupt or rapid turning of the flow during diffusion. The attendant strong shock-boundary-layer interactions can result in separation of the flow from the walls of the duct with large mixing losses At Mach numbers of 2.0 and higher, to obtain high internal performance *it is generally necessary to incorporate* a means of boundary-layer control or removal in the vicinity of the throat Such devices include bleed perforations, flush slots, or ram scoops.

When applied to various jet engines, an efficient supersonic diffuser, or inlet, besides attaining high internal performance in terms of total-pressure recovery or kinetic-energy efficiency, must also have a low external drag Total drag consists of (1) the pressure drag acting on the external cowl projected frontal area, (2) spillage or additive drag occurring when the inlet spills flow around the cowl (defined as the pressure integral along the entering streamline), and (3) friction drag on the external cowl surface.

From considerations of maximum airplane range or acceleration, the trade-off between pressure re-

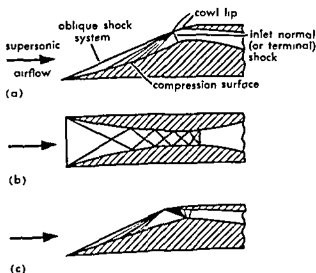
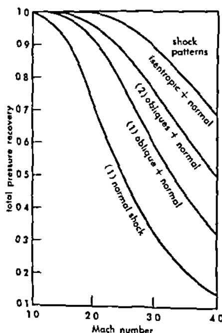


Fig 1. Basic compression systems (a) External compression. (b) Internal compression. (c) External and internal compression.



2 Theoretical performance comparison of external-compression inlets

very and drag can be evaluated to determine an optimum on-design inlet. The relative importance of pressure recovery and drag changes with flight speed. In general, recovery predominates at the low supersonic Mach numbers, drag at Mach numbers above about 3.5.

The supersonic diffuser in particular installations must meet the airflow schedule dictated by the particular type of power plant throughout its entire operating range of flight speed and altitude with the best combination of recovery and drag. During off-design conditions, the inlet handles excess airflow. Most often the inlet is sized for the high-speed cruise, or design, condition, thus making it oversized for less than design speeds. The penalties associated with various excess-air handling methods can wipe out the entire thrust margin of the airplane, unless sufficient care is exercised. Techniques that are available include inlet-shock spillage, bypass through an auxiliary exit, or bypass through an ejector exhaust nozzle.

The physical geometry of the supersonic diffuser will generally depend upon installation and airplane configuration. For ex... ine ne... inlet... where ... is submerged in the fuselage, might best use a two-dimensional induction system. In general, the two-dimensional configurations are more adaptable to the variable-geometry requirements of engine inlet matching; their axisymmetric counterparts possess some inherent structural hoop-tension advantages. [J.F.CO.; L.J.O.]

Bibliography: J. F. Connors and C. R. Meyer, *Design Criteria for Axisymmetric and Two-Dimensional Supersonic Inlets and Exits*, Natl Advisory Comm. Aeronaut., Tech. Note 3589, 1956; J. C. Evvard and J. W. Blakey, *The Use of Perforated Inlets for Efficient Supersonic Diffusion*, Natl Advisory Comm. Aeronaut., RM E51B10, 1951; D. P. Hearth and J. F. Connors, *A Performance Analysis of Methods for Handling Excess Inlet Flow at Supersonic Speeds*, Natl. Advisory Comm. Aeronaut., Tech. Note 4270, 1958; A. Kantrowitz and C. Donaldson, *Preliminary Investigation of Supersonic Diffusers*, Natl. Advisory Comm. Aeronaut., WR L-713, 1945; R. W. Luidens and R. J. Weber, *A Method for Evaluating Jet-Propulsion-System Components in Terms of Missile Performance*, IAS Preprint No. 656, 1956; M. Sibulkin, *Theoretical and Experimental Investigation of Additive Drag*, Natl Advisory Comm. Aeronaut., Rept. 1187, 1954.

Supersonic flight

Relative motion of a solid body and a gas at a velocity greater than that of sound propagation under the same conditions. The general characteristics of supersonic flight can be understood by considering the laws of propagation of a disturbance, or pressure impulse, in a compressible fluid.

Formation of Mach cone. If the fluid is at rest, the pressure impulse propagates uniformly with the velocity of sound in all directions, the effect always acting along an ever-increasing spherical surface. If, however, the source of the impulse is placed in a uniform stream, the impulse will be carried by the stream simultaneously with its propagation at sonic velocity relative to the stream. Hence the resulting propagation is faster in the direction of the stream and slower against the stream. If the stream velocity equals the velocity of sound, the effect of the impulse cannot reach every point in space, but is restricted to the half-space bounded by a plane perpendicular to the flow direction. The source of the impulse is no longer able to send signals upstream.

If the velocity of the stream past the source of disturbance is supersonic, the effect of the impulse is restricted to a cone whose vertex is the source of the impulse and whose vertex angle decreases from 90° (corresponding to Mach number equal to 1) to smaller and smaller values as the Mach num-

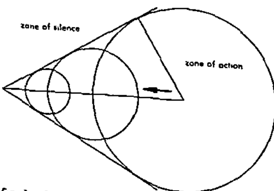


Fig. 1. Generation of Mach wave by body moving at supersonic velocity

ber of the stream increases (Fig. 1). In fact, the trigonometric sine of the vertex half-angle is equal to the reciprocal of the Mach number. This angle is called the Mach angle and the cone so formed, the Mach cone. If the source of the pressure impulse travels through the air at rest, the conditions are analogous. *See* MACH NUMBER.

The cone described above separates a zone of action from a zone of silence. Here lies a fundamental difference between subsonic and supersonic motion of a body. In subsonic motion, the effect of the disturbance, although decreasing with distance, reaches every point of the space surrounding the body. In supersonic motion, the disturbance caused by the body is restricted to the inside of the Mach cone. For example, if a projectile passes over an observer's head at supersonic speed, he hears it only after it has passed.

General rules. These considerations can be formalized into three general rules: (1) the rule of forbidden signals, (2) zone of action and zone of silence, and (3) rule of concentrated action. Because a slight pressure change is propagated with sound velocity, the effect of pressure changes produced in the air by a body moving at a speed faster than sound cannot reach points ahead of the body; this is the rule of forbidden signals. The zone of action and the zone of silence arise because a stationary point source in a supersonic stream produces action only at points that lie on or inside the Mach cone extending downstream of the point source. Conversely, the pressure and velocity at an arbitrary point in the stream can be influenced only by disturbances acting at points that lie on or inside a cone of the same vertex angle extending upstream of the point considered. The rule of concentrated action thus follows For a body moving at supersonic speed, the major portion of the effect is concentrated in the vicinity of the Mach cone that forms the outer limit of the zone of action.

Shock wave. Consider the supersonic motion of a wing moving into air at rest. Because signals cannot propagate ahead of the wing, the presence of the wing has no effect on the undisturbed air until the wing passes through it. Hence there must be an *abrupt change in the properties of the undisturbed air* as it begins to flow over the wing. This

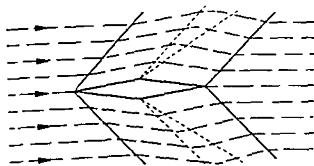


Fig. 2. Streamlines, shock waves, and expansions around a supersonic wing.

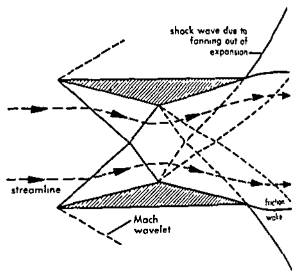


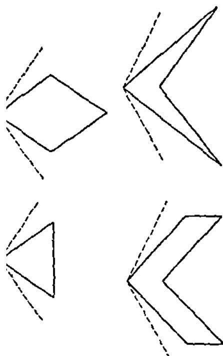
Fig. 3. Flow around Busemann biplane

abrupt change takes place in a shock wave which is attached to the leading edge of the wing, provided that the leading edge is sharp and the flight Mach number is sufficiently large As the air passes through the shock wave its pressure, temperature, and density are markedly increased.

Further aft of the leading edge, the pressure of the air is decreased as the air expands over the surface of the wing. Hence the pressure acting on the front part of the wing is higher than the ambient pressure, and the pressure acting on the rear part of the wing is lower than the ambient pressure. The pressure difference between front and rear parts produces a drag, even in the absence of skin friction and flow separation. The wing produces a system of compression and expansion waves which move with it (Fig. 2).

Wave drag. The work which must be done to create and carry these waves is the work which must be done to overcome the drag. This phenomenon is similar to that of a speedboat moving with a velocity greater than the velocity of the surface waves In this case, the boat carries the waves that it produces, and the work done is a large part of the total resistance of the boat Because of this analogy, supersonic drag is called wave drag It is peculiar to supersonic flight, and it may represent *the major portion of the total drag* of a body (Drag associated with skin friction and flow separation is still present in supersonic flight.) A detailed consideration of the wave drag leads to the conclusion that supersonic wings should have sharp leading and trailing edges, and the thickness ratio of the wing should be as small as possible to reduce the wave drag *See* AERODYNAMIC WAVE DRAG.

Lift and induced drag. The theory of lift of an aircraft wing moving at subsonic speeds is based on the concept of circulation In the case of supersonic flight, the same concept can be used, but the flow pattern associated with the circulation must be restricted to the surface and interior of the Mach cone whose vertex lies at the point of application of the lift force. Induced drag also exists in the supersonic case as the subsonic



3 4 Types of finite supersonic sweptback wing

se However, in the supersonic case, the production of lift requires, in addition to the induced drag, certain amount of wave drag corresponding to the energy radiated to infinity along the Mach cone. This drag is found to be proportional to the square of the lift produced, that is, it follows a law similar to that governing induced drag in subsonic flight.

Planform. Because the wave drag is proportional to the square of the thickness ratio of the wing and the square of the lift produced, small thickness ratios and small lift coefficients are favored. However, this generalization is limited by requirements of weight and strength of construction. Hence, other methods for reducing wave drag are sought.

A. Busemann has proposed the use of a biplane which is designed in such a way that the waves emanating from the upper wing are compensated, or cancelled, by the lower wing (Fig 3). In this scheme, there are no waves external to the lifting surface so that no energy can leave the moving system, hence there can be no wave drag. The idea essentially uses the concept of wave interference to reduce wave drag.

Another idea involves the sweptback wing (Fig). Because the wing characteristics depend on the component of velocity normal to the leading edge, it may be possible to provide sufficient sweepback that the flow velocity normal to the wing axis becomes subsonic. In this case, there can be no wave drag. Because in supersonic flight the unavoidable profile drag is relatively large compared to induced drag, a small aspect ratio is acceptable. Hence the use of a delta wing appears promising; it can take advantage of both a large sweepback angle and a small thickness ratio. Furthermore, the displacement of the center of pressure in transition from subsonic to supersonic flight is smaller than for more conventional planforms.

Effect on range. The range of an aircraft depends, among other things, on the lift-drag ratio of the aircraft. In the case of supersonic airplanes, the aerodynamic efficiency depends to a great extent on effective use of volume, such as the ability of the designer to provide volume for fuel or propellants. Consequently, the supersonic aircraft becomes a large body with small winglets, and the lift-drag ratio becomes unfavorable.

A simple approximate analysis shows that the lift-drag ratio of such an aircraft depends essentially on the ratio of the local ambient pressure to the weight of the fuselage per square foot of frontal area. Hence the range of a supersonic aircraft is greatly enhanced by flight at high altitude and by denser loading or by increasing the over-all dimensions of the aircraft. These conclusions again show a major difference between subsonic and supersonic flight, because for subsonic flight altitude has no significant effect on the range, and velocity and size have only secondary effect. *See* HYPERSONIC FLIGHT; SUBSONIC FLIGHT; TRANSONIC FLIGHT [J.E.SC.]

Bibliography: E. A. Bonney, *Engineering Supersonic Aerodynamics*, 1950, C. duP. Donaldson, J. V. Charyk, and M. Summerfield (eds.), *High Speed Aerodynamics and Jet Propulsion*, vol. 6, 1954, A. Ferri, *Elements of Aerodynamics of Supersonic Flows*, 1949; H. W. Liepmann and A. Roshko, *Elements of Gasdynamics*, 1957; T. Von Kármán, *Aerodynamics*, 1954; T. Von Kármán, Supersonic aerodynamics—principles and applications, *J Aeronaut Sci*, 14(7):373–409, 1947.

Suppression

The elimination of some undesired component of a signal. In automobile radio installations, suppression techniques are essential to prevent ignition interference from reaching the radio circuits. Radio receivers themselves often contain special noise-suppression circuits and devices; this is particularly true for communication receivers that operate in cro... ded ...

a

tl

C ...ppression technique involves reducing the receiver gain suddenly after each high-power pulse is transmitted, then gradually and automatically restoring normal gain so that nearby echo signals are amplified much less than the desired distant echo signals. *See* NOISE, ELECTRICAL. [J MR]

Suppressor

A device used to reduce or eliminate noise or other signals that interfere with the operation of a communication system. This term may be applied to a noise filter in a radio receiver (see NOISE FILTER, RADIO), but it is more frequently used to describe a device applied at the noise source, such as a re-

sistor used in series with spark plugs of a gasoline engine, or a capacitor across the terminals of a commutator motor or other sparking device that acts as a noise generator. The term suppressor may also be applied to a filter used in power leads of an electronic device to eliminate unwanted signals or noise. *See* FILTER, ELECTRIC. [W.R.L.]

Surface and solid of revolution

A surface of revolution is generated by revolving a plane curve about a line in its plane. The generating curve may consist of one or more arcs or line segments connected together. A solid of revolution is generated by revolving a connected plane region about a line in its plane not cutting the region. The boundary of a solid of revolution is a surface of revolution. The line about which the generating plane curve or region is revolved is called the axis of revolution, or simply the axis.

Generation of surfaces. In euclidean solid geometry the words cylinder, cone, and sphere commonly refer to solids of revolution, and the corresponding surfaces of revolution are called a cylindrical surface, a conical surface, or a spherical surface. However, in analytic geometry and more generally in modern usage, the words cylinder, cone, and sphere refer to surfaces, not solids. A circular cylinder is a surface generated by revolving an infinite line about a parallel axis, and a circular cone is a surface generated by revolving an infinite line about an intersecting axis. A spheroid is a surface obtained by revolving an ellipse about one of its axes. Paraboloids and hyperboloids of revolution are obtained by revolving parabolas and hyperbolas about their axes.

Two surfaces of revolution that play a special role as interesting examples in differential geometry are those obtained by revolving a catenary or a tractrix about its axis. Surface tension pulls a soap film spread on a wire boundary to form a so-called minimal surface, having a smaller area than any nearby surface with the same boundary. If the boundary consists of two circles with centers on a common axis perpendicular to their planes, and not too far apart, the minimal surface connecting them is part of a catenoid of revolution, obtained by revolving about the x axis the curve

$$y = a \cosh (x/a)$$

in which a uniform inextensible chain hangs under its own weight. As the circles are pulled apart there comes a point at which the soap film breaks in two and forms two separated plane circular disks. *See* CATENARY; TRACTRIX.

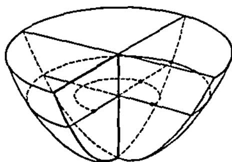


Fig. 1. Paraboloid of revolution.

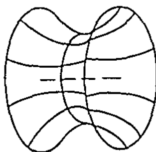


Fig. 2. Catenoid of revolution.

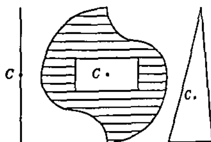


Fig. 3. Centroids, C.

The tractrix of revolution is of interest because it is a surface of constant negative curvature on which the geometry is noneuclidean. The neighborhoods of two points are congruent to each other. But if three points on the surface are joined on the surface by curves of shortest length called "lines," the sum of the angles in the resulting triangle is always less than 180° and through a given point there is more than one "parallel" to a given line not through the given point.

Theorems of Pappus. Two important theorems that bear the name of Pappus of Alexandria, who lived in the third century A.D., give formulas for computing respectively (1) the volume V of a solid of revolution generated by revolving a connected plane region of finite area A about an axis in its plane not cutting the region; and (2) the area S of the surface of revolution obtained by revolving a simple plane curve of finite length l about an axis in its plane, but not cutting the curve. Both theorems involve the notion of the centroid or center of gravity, of the plane region or curve. If this centroid C is at a distance c from the axis, it moves a distance $2\pi c$ during the rotation that generates the solid or surface.

Theorem I. The volume V of the solid generated by revolving a connected plane region about an axis in its plane, but not cutting the region, is equal to the product of the area A of the generating region times the circumference $2\pi c$ of the circular path through which the centroid moves.

Volume of solid of rev [illegible] $V = 2\pi cA$

Theorem II. The area S of the surface generated ·revolving a connected plane curve about an axis its plane, but not cutting the curve, is equal to e product of the arc length l of the generating rve times the circumference $2\pi c$ of the circular th through which the centroid moves.

rea of surface of revolution $\quad S = 2\pi cl$

The centroid C of a plane region (or curve) ving a center of symmetry is that center Intui- ely the centroid of a connected plane region or rve not having a center of symmetry may be ought of as follows Let a model of the region or rve be cut out of thin sheet metal or wire of uni- rm thickness. Then the centroid is a point C in ane of the region or curve such that, if the model placed horizontally over a knife-edge supporting

a knife-edge along any median so its centroid the point of intersection of the medians.

Formulas for obtaining the coordinates $(\bar{x},\bar{y})$ of e centroid of any plane region or curve are derived the calculus·

entroid $(\bar{x},\bar{y})$ of area

$$\bar{x}A = \iint x\,dx\,dy \qquad \bar{y}A = \iint y\,dx\,dy$$

entroid $(\bar{x}\bar{y})$ of curve

$$\bar{x}l = \int x\,ds \qquad \bar{y}l = \int y\,ds$$

orresponding formulas for volumes and surfaces revolution obtained by revolving about the x is a plane region of area A or a simple curve of ngth l lying wholly above the x axis are

$$\text{Volume} = 2\pi \iint y\,dx\,dy$$
$$\text{Surface} = 2\pi \int y\,ds$$

e CENTROIDS OF AREAS AND LINES

Three examples will illustrate Pappus' theo- ms

Example 1 Find the volume of a right circular one by Pappus' theorem Solution If a right tri- gle AOB is revolved about its side OA, it gener- es a solid right circular cone of height $h = \overline{OA}$ nd radius of base $r = \overline{OB}$ The area of the plane gion is $A = rh/2$ and the centroid C is at the stance $c = r/3$ from the axis OA Hence the vol- me is given by

$$V = 2\pi(r/3)(rh/2) = \pi r^2h/3$$

Example 2 Find the distance c from the centroid f a semicircular arc of radius r to the center of the rcle Solution. The arc length is $l = \pi r$, and the rea S of the surface of the sphere generated by volving the semicircle about its diameter is $S = 4\pi r^2$ Hence the centroid distance is

$$c = S/2\pi l = 4\pi r^2/2\pi^2 r = 2r/\pi$$

Example 3 Find the volume and surface area of solid torus obtained by revolving a circle of ra-

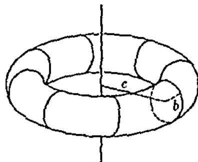


Fig 4 Torus.

dius b about an axis in its plane at distance $c > b$ from the center. Solution:

$$V = 2\pi c(\pi b^2) = 2\pi^2 b^2 c \quad \text{(volume of torus)}$$
$$S = 2\pi c(2\pi b) = 4\pi^2 bc \quad \text{(surface of torus)}$$

See CONE; CYLINDER, ELLIPSOID AND SPHEROID; HYPERBOLOID, PARABOLOID; QUADRIC SURFACE; SPHEROID; TORUS [J.S.F.]

Surface coating

A substance applied to other materials to change the surface properties, such as color, gloss, resist- ance to wear or chemical attack, or permeability, without changing the bulk properties. The term in- cludes such materials as ... and lacquers. ... based on a ve... after being spread out in a relatively thin film, changes to a solid. This change, called drying, may be due entirely to the evaporation of a volatile sol- vent, or it may be caused by a chemical reaction, such as oxidation and polymerization Opaque ma- terials called pigments, dispersed in the vehicle, contribute color, opacity, and increased durabil- ity and resistance.

Paints are organic coatings in which the vehicle is usually a drying oil, such as linseed, soybean, tung, or dehydrated castor, or a varnish The term paint is often used to cover all organic coatings. Enamels are paints in which the pigment is very well dispersed and the vehicle is chosen to flow out to an extremely smooth finish, usually with a high gloss, although semigloss enamels are also made Varnishes are clear, unpigmented coatings, made by dissolving a resin, or the reaction prod- ucts of a resin and a drying oil, in a suitable sol- vent. Lacquers are coatings for which the ...

... used for a specific purpose by the name of the products most commonly used for that purpose, even though it does not otherwise meet the defini- tion For example, nearly all clear coatings for metal are called lacquers, regardless of their com- position

Organic coatings are usually referred to as deco- rative or protective, depending upon whether the primary reason for their use is to change (or pre-

serve) the appearance or to protect the surface. Often both purposes are involved.

The origin of organic coatings is lost in ancient history. Pictures on cave walls indicate that paints were known to prehistoric man, and the use of bituminous materials and pitch for waterproofing and preservation is mentioned in the Bible and was known to the ancient Egyptians. Paints were also used for the decoration of furniture and for the production of pictures in ancient times. More recently, the use of drying oils, rosin, and other resins, as well as the development of inexpensive pigments suitable for use in paints, reduced the cost to the point that houses and other structures could be painted. The wide use of iron and steel during the past century made the use of organic coatings essential to reduce losses from corrosion and chemical attack.

As more materials of construction become available, the need to modify the surface in some way continues to grow. Even materials which are resistant to most forms of attack, such as stainless steel, may need protection from certain forces. Many surfaces which are widely believed to be satisfactory without coatings are actually attacked by weather, salt, atmospheric polution, or other factors, and must be protected for certain uses.

The wide variety of surfaces which must be protected and decorated has given rise to an infinite variety of coatings. Many of these are not called paints, although a large number fall under almost any definition of paint. Mastic coatings are usually similar to paints in composition, but are applied *at a heavier consistency, usually with a trowel.* Other organic coatings include sheet plastics, which may be formed in place by a doctor blade, formed first and applied with an adhesive, or applied as a dry powder or a dispersion and fixed by heating above the melting point. Coatings of this type are widely used for sheet metals and for textiles. Inorganic surface coatings include vitreous enamels, metal plating, and metallic salts deposited from solution. This last method is widely used to prepare metal surfaces for subsequent painting *See* DRIER (PAINT); DRYING OIL; ELECTROPLATING OF METALS; ENAMEL, NONVITREOUS; FILLER; JAPANNING; LACQUFR; METAL COATINGS; PAINT; PIGMENT; PRIMER (SURFACE COATING); SHELLAC, THINNER; VARNISH. [F SD.]

Surface condenser

A heat transfer device used to condense a vapor, usually steam, by absorbing its latent heat in a cooling fluid. ordinarily water. Most surface condensers consist of a chamber containing a large number of 0.5- to 1-in. diameter corrosion-resisting alloy tubes through which cooling water flows. The vapor contacts the outside surface of the tubes and is condensed on them The tubes are arranged so that the cooling water passes through the vapor space one or more times. About 90% of the surface is used for condensing vapor and the remaining 10% for cooling noncondensable gases. Air coolers are normally an integral part of the condenser but may be separate and external to it. The condensate is removed by a condensate pump and the noncondensables by a vacuum pump *See* CONDENSER, VAPOR; STEAM CONDENSER. [J T.S.]

Surface hardening of steel

When properly carried out, surface hardening produces high wear resistance, great strength, and good impact resistance to meet the most exacting engineering requirements.

Gears are among the parts most commonly surface hardened, and millions of these are hardened each month in the automotive industry alone. Other parts surface hardened in large quantities are ball and roller bearings, piston pins, camshafts, tappets, and spindles.

The case-hardening processes described in this article produce surface layers on a steel alloy that are substantially harder than the interior. The surface layer is called the case and the interior is referred to as the core.

The processes common to case hardening are (1) carburizing, (2) cyaniding, (3) carbonitriding, (4) nitriding, (5) induction hardening, and (6) flame hardening Each of these processes has its place in industry, and is best used under certain conditions of fabrication and with certain engineering requirements.

For other methods of surface hardening such as chromium plating, metallizing, hard metal overlaying, calorizing, chromizing, and chemical coatings *see* ELECTROPLATING OF METALS; HEAT TREATMENT (METALS AND ALLOYS); METAL COATINGS.

Carburizing. This is a process whereby carbon is introduced into solid steel above its transformation temperature in contact with a carbon-containing medium which may be either solid, liquid, or gaseous. When the carbon-containing medium is solid, the process is called pack carburizing; when liquid, it is called liquid-salt carburizing; and when gaseous, it is called gas carburizing.

Pack carburizing. Pack carburizing, or as it is often termed box carburizing, consists of packing the steel to be carburized in a box and surrounding it with charcoal which has been treated with a so-called energizer. This energizer usually consists of an alkali or metal carbonate such as *barium carbonate*. The carburizing compound should contain not less than about 10% of the energizer to accomplish the best rate of carburizing.

After packing, the box is placed in a furnace and brought to a temperature of 1650–1700°F and allowed to remain in the furnace long enough to give the required depth of case. The usual time will be 4–20 hours. The usual case depth will be from 0.025 to 0.045 in., although depths up to 0.125 in and more are not uncommon.

After carburizing, the parts may be hardened by removing them from the box while they are still at or near carburizing temperature and quenching directly in a cooling medium, usually oil Alternatively, the parts may be permitted to cool in the

arburizing box to room temperature, removed

as the automotive ring gear, which is oil ched in a press in which the gear is held firmly die so that there is very little distortion. Most , however, are quenched directly from the car- ing heat, and if made from the proper steel, have maximum resistance to fatigue failure. direct quench method has the lowest cost. me of the steels used for surface hardening listed in the table. Those used for carburizing

age composition of steels used in surface hardening

signation	Carbon, %	Manganese, %	Nickel, %	Chromium, %	Molybdenum, %
1019	0 17	0 85			
1022	0 20	0 85			
3310	0 10	0 52	3 50	1 55	
E 4118	0 20	0 80		0 50	0 12
E 4320	0 20	0 55	1 80	0 80	0 25
E 4310	0 40	0 70	1 80	0 80	0 25
E 4620	0 20	0 55	1 80		0 25
E 8620	0 20	0 80	0 55	0 50	0 20
E 1035	0 35	0 75			
tralloy (aluminum, 1 05%)	0 37	0 55		1 60	0 35

e Society of Automotive Engineers (SAE) designations 1019, 1022, 3310, 4118, 4320, 4620, and 620 These steels are specified to be fine grained o that when quenched from 1700°F, the core of he steel will be fine grained Steels 1019 and 1022 re plain carbon and must be quenched in water or rine to obtain proper hardness Such drastic treatment usually causes excessive distortion and can e applied only when this is permissible The plain arbon steels, however, are the cheapest

Parts in which low distortion is required are made from alloy steel which will harden when quenched in oil Larger parts cool more slowly when quenched, and thus require more alloy in the steel to harden properly The carburizing steels in the table which have the greatest ability to harden, that is have the highest hardenability, are SAE 3310 and 4320

Liquid salt carburizing Carburizing may be carried on in a hot liquid bath by immersing the steel in the bath and then quenching in cold water or oil. The bath usually contains appreciable quantities of barium or calcium salts (chlorides) and 8 23% of sodium cyanide The carburizing is carried on at 1550–1750°F. This method is used for light case depths of 0 005–0 030 in The higher the temperature of operation, the faster the carburizing rate but the more trouble from pot failures and salt vaporization

Gas carburizing In this process, the parts are exposed to an atmosphere containing gaseous hydrocarbons or carbon monoxide at the carburizing temperature Since it is not necessary to heat through carburizing boxes and charcoal, as in box carburizing, gas carburizing can give the same case depth in less time Gases used in carburizing may be natural gas, commercial manufactured gas, butane, propane, or methane. These gases are usually diluted to give easier control of the process.

The control of surface carbon content and better control of case depth, usually coupled with a somewhat lower cost, are the advantages of the gas-carburizing process

The subsequent hardening, tempering, measurement of case depth, and hardness are the same as for pack carburizing.

Tempering. It is considered good practice to reheat the quenched parts to about 275–400°F and hold at these temperatures for 1–3 hours or more to relieve the residual stresses and improve resistance to cracking during grinding. Frequently, the tempering operation is not carried out, and the omission is permissible with many parts

Case depth. The depth of the carburized or hardened portion is called the case depth. Effective case depth is defined as the distance (measured perpendicularly) from the surface of a hardened case to a point equivalent to a hardness of Rockwell C50 Total case depth is defined as the distance from the surface of the hardened or unhardened case to a point where differences in chemical or mechanical properties of the case and core no longer can be distinguished

It is difficult to control the total case depth to limits closer than 0 010 in total spread from maximum to minimum in the pack-carburizing process. This process should not be used in industrial work when total case depths of less than 0 020 in. are desired Surface hardness of Rockwell C60 minimum is obtainable with this process, although in many applications, such as automotive gears, a minimum of C57 is acceptable

Cyaniding. In the cyaniding process, a thin, hard case is produced, usually up to 0 008 in., although much thinner cases in the range of 0 001–0 003 in. are most common In this process, the salts used are combinations of sodium chloride, sodium carbonate, and sodium cyanide The latter is used in the range of 15–40%. This process is usually carried on at temperatures of 1350–1550°F and 15–30 min in the salt is common. Quenching in cold water, brine, or oil direct from the cyaniding pot is necessary to obtain good hardness

Both carburizing steels and medium carbon steels such as SAE 1035 are often cyanided The hardness obtainable by this process will resist the action of a Rockwell C60 file. Cyanided case and to some extent the liquid-carburized case contains nitrides as well as carbides of iron to produce good hardness and wear resistance.

Carbonitriding. This process has recently come into favor, and is being used to a large extent instead of cyaniding since it eliminates all the troubles with liquid salts and pot failures. The steel is heated in a gas-carburizing atmosphere to which ammonia gas is added This imparts nitrogen as well as carbon into the surface and gives a file-hard

surface with oil quenching, even with plain carbon steels. About 1% of ammonia in the carburizing atmosphere imparts this property.

Nitriding. Nitriding is a process in which a steel alloy, usually of special composition, is surface hardened without quenching by heating in an atmosphere of ammonia or in contact with a nitrogen-containing material so that nitrogen is absorbed. The special alloy steels used for nitriding are known as nitralloy, and contain aluminum as one of the alloying elements.

Before nitriding, the steel is oil-quenched and tempered to 1100–1300°F, machined, and then retempered at these temperatures to remove the machining strains. Any decarburized areas must be removed, since these areas will produce very brittle surfaces after nitriding.

Nitriding is accomplished by heating the steel in an atmosphere of dry ammonia at about 900–1000°F for a period of 24–72 hours. Fifty hours at 970°F is a common practice and produces a case depth of about 0 020 in. Very high hardness results from proper nitriding, which gives a range of 925–1400 diamond pyramid hardness number, DPH. This hardness is beyond the Rockwell C scale; 940 DPH corresponds to Rockwell C68 Thus, it is evident that nitriding can be used to develop much higher hardness than carburizing can Also, in general, there is less distortion than is encountered during carburizing and hardening, because nitriding is accomplished at much lower temperatures, and because it is not necessary to quench the steel after nitriding. Nitriding, however, causes growth of the steel which results in some distortion. Proper allowance can be made for this growth in dimensioning the fabricated part before nitriding Because the extreme outer nitrided surface has a tendency to be very brittle, it is customary to grind 0 001–0.003 in off this surface after nitriding.

It is possible to nitride some of the common alloy steels which do not contain appreciable amounts of aluminum, although the hardness in this case is much lower than with nitralloy (400–600 DPH). When SAE 4310 steel is so treated, it shows improved resistance to wear, increased endurance limit, and nonseizing properties superior to those when not nitrided Another advantage of nitrided cases is that they retain their room temperature hardness up to 750°F. This is not true of carburized cases.

Induction hardening. Induction hardening is a process of surface heating a steel alloy above the [illegible]

ing a high-frequency current which passes through a coil that is around the section to be hardened The frequency usually varies from 1000–10,000 cycles per minute (cpm) although frequencies above 500,000 cpm have been used. The instantaneous-heating effect varies inversely as the square root of the frequency. Thus, 9 kilocycles (kc) heats instantly 5 times deeper than 225 kc The heating time in this process is usually a few seconds, 3–10 sec is quite common, and the part is cooled by water, oil, air, or by self cooling.

The steels used in this process are usually plain medium-carbon steels in the range of 0 35–0 50% carbon; others are alloy constructional steels, hardenable stainless steels, cast iron, and malleable iron.

This process permits convenient hardening of any portion of a part, gives low distortion, produces favorable residual stresses so that the part will have excellent resistance to fatigue stressing and gives good reproducibility of the hardness pattern. The effective case depths usually vary from 0 025 to 0.20 in., but may vary over a wider range if desired. Parts commonly hardened are axle shafts crankshafts, and camshafts.

The installation for induction hardening equipment is rather costly so that it is best suited for high-production rates.

Flame hardening. This may be defined as a process of heating the surface layer of steel or other hardenable alloy above the transformation range by means of a high-temperature flame and quenching The time of heating is short, and an effective case depth of 1/16–1/8 in. is usually obtained The heating time for average operations is 30–40 sec The quenching is usually done with water or compressed air

Although any type of hardenable steel may be used, a range of 0.35–0 50% carbon is most common. Cast iron is also often flame hardened

Flame hardening has practically the same advantages as induction hardening. With flame hardening, the reading of the surface temperature is difficult because of interference from the light of the flame Much care is required to ensure good reproducibility Parts frequently hardened by this method are camshafts, large parts such as lathe ways, and large gears *See* METAL, MECHANICAL PROPERTIES OF; METAL FORMING; STAINLESS STEEL, STEEL. [W E.J O]

Surface phenomenon

One of the numerous reactions and interactions which occur at the surfaces of solids and liquids as opposed to those which occur within a solid or a liquid as a whole. A more correct terminology would be interfacial phenomena, since a surface represents a boundary between at least two coexistent phases, that is, an interface between a solid or a liquid and another solid, liquid, or gas. Surface and interface are often used interchangeably.

Some characteristic phenomena associated with surfaces are detectable at all liquid or solid surfaces Others become apparent only in systems that have a large ratio of surface area to volume. The extent of surface area thus distinguishes two general categories of surface phenomena. However, both types owe their existence to such characteristic properties of surfaces as surface potential and surface energy, or surface [illegible]

Some properties of surfaces which can be measured on a macroscopic scale are wettability, adhesion, friction, and thermionic and photoelectric emission, as well as surface or interfacial tension. The characteristic magnitudes of these various surface properties are highly dependent on the cleanliness of the surface; great changes *in their values* can be brought about by the introduction (deliberate or otherwise) of foreign materials Practical use is often made of this, for example, in the reduction of friction by *lubrication and* in the process of waterproofing, where wettability of a surface *is reduced by* the application of a thin film of a

lower the surface tension to a marked degree when added in small amounts are known as surface-active agents, or surfactants Solutions of such substances have remarkable powers of wetting, as indicated by the success of synthetic detergents in cleansing processes The diversity of processes dependent on macroscopic surface properties is illustrated by such examples as dyeing, soldering, *mineral* flotation, and the restriction of water evaporation from open reservoirs. In this last example an insoluble monolayer of some appropriate substance, such as cetyl alcohol is allowed to spread spontaneously on the water surface

There are innumerable phenomena which become apparent upon the development of the high *ratio* of surface area to volume. To illustrate the effect of increased surface area on the surface energy of a system, consider a cube of 1-cm edge which has a surface area of 6 cm², and a surface energy *of* 6 γ (where γ is the surface tension in dynes/cm or ergs/cm²) If the cube is divided *into* 10^{18} *cubes* of edge 100 A (10^{-6} cm), the surface area becomes 6 000 000 cm² and the surface energy becomes $6 \times 10^6\ \gamma$ This enormous increase of *energy affects*, among other things, solubility, rate of *solubility*, chemical activity and color

Perhaps the most significant of the properties which can be detected *in the* presence of a large surface area is that of adsorption. The ability of solid and liquid surfaces to attract and hold *re*-tet with, or cause to react with each other the molecules of gases, vapors, liquids, and even ions is *of great importance* Contact catalysis, soil nutrition, ion exchange, wine purification and chromatographic separation are *processes* dependent on the phenomenon of adsorption Other manifestations of the high *ratio* of surface area to volume are the colloidal systems *See* ADSORPTION, COLLOID, FLOTATION, INTERFACE OF PHASES, LUBRICANT, SOAP AND DETERGENT SURFACE ACTIVE AGENT, SURFACE TENSION [A L D, W O M]

Surface tension

The force acting in the surface of a liquid, tending to minimize the area of the surface Surface forces, or more generally, interfacial forces govern such phenomena as the wetting or nonwetting of solids by liquids, the capillary rise of liquids in fine tubes and wicks, and the curvature of free-liquid surfaces. The action of detergents and antifrothing agents and the flotation separation of minerals depend *upon* the surface tensions of liquids.

Surface energy. In the body of a liquid, the time-averaged force exerted on any given molecule by its neighbors is zero Even though such a molecule may undergo diffusive displacements because of random collisions with other molecules, there exist no directed forces upon it of long duration. It is equally likely to be momentarily displaced in one direction as in any other. In the surface of a liquid, the situation is quite different; beyond *the free* surface, there exist no molecules to counteract the forces of attraction exerted by molecules *in the* interior for molecules in the surface. In consequence, molecules in the surface of a liquid experience a net attraction toward the interior of a drop. These centrally directed *forces cause the droplet to* assume a *spherical shape, thereby* minimizing both the free energy and surface area.

From the macroscopic point of view, surface tension may be regarded either as a force exerted normally to a unit length in the surface, or as the *work which must* be expended upon the liquid to *increase its* area by unity Accordingly, surface tension is expressed in cgs units of dynes/cm or ergs/cm². From the microscopic point of view, the surface tension (or its equivalent, surface energy) is the reversible isothermal work which must be done in bringing molecules from the interior of the liquid to the surface and creating 1 cm² of new surface thereby.

Most liquids have surface tensions of 20–40 dynes/cm at room temperature, but water has the exceptionally high value of 72.75 dynes/cm at 20°C Condensed gases such as helium and nitrogen have quite low surface tensions (0 098 dynes/cm at 4 3°K and 6 2 dynes/cm *at* 90.2°K, respectively) Liquid metals have large surface tensions by comparison, mercury, 470 dynes/cm; and liquid copper at 1131°C has a surface tension of 1103 dynes/cm in hydrogen gas. Small but significant differences in the surface *tensions of liquids* depend upon the composition of the vapor *phase*

In the wetting or nonwetting of solids by liquids, the criterion *employed is* the contact angle between the solid *and the* liquid (measured through the liquid), (Fig 1). A liquid is said *to* wet a *solid* if the contact angle θ lies between 0° and 90°, and not to wet the solid if the contact angle

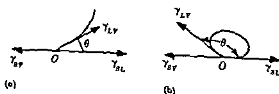


Fig 1 Contact angle (a) Liquid wetting solid. (b) Liquid not wetting solid

lies between 90° and 180°. Three interfaces exist when a droplet of liquid contacts a solid, and three corresponding interfacial tensions exist: γ_{SL}, γ_{SV}, and γ_{LV}. The subscripts S, L, and V refer to solid, liquid, and vapor. At equilibrium, a balance of interfacial tensions exists at the line of common contact, which intersects the figures at point O. For the case of a liquid which wets the solid ($\theta < 90°$), this equilibrium is expressed by the relation:

$$\gamma_{SV} = \gamma_{SL} + \gamma_{LV} \cos \theta$$

Capillarity. Liquids which wet the walls of fine capillary tubes rise to a height which depends upon the tube radius, the surface tension, the liquid density, and the contact angle. In Fig. 2, a liquid of density ρ is shown as having risen to a height h in a capillary whose radius is r. A balance exists between the force exerted by gravity on the mass of liquid raised in the capillary and the opposing force caused by surface tension. The former is $\pi r^2 h \rho g$, whereas the latter is $2\pi r \gamma$, assuming the contact angle to be zero. It is clear that $h = 2\gamma/r\rho g$, and that the capillary rise varies inversely with the tube radius and the liquid density. Liquids which do not wet the capillary walls are depressed in height according to the same equation.

The shape of the free surface of a liquid in a vessel is only an approximation to a plane In narrow tubes the meniscus of a liquid is concave upward if the liquid wets the tube, and conversely convex upward if it does not wet the tube A pressure difference exists between the concave and convex sides of the surface, the excess pressure on the concave side over the convex side being given by the relation

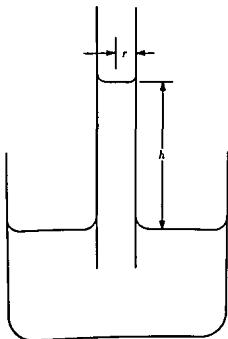


Fig. 2. Rise of liquid in capillary tube.

$$p = \frac{1}{r_1} + \frac{1}{r_2}$$

where r_1 and r_2 are the principal radii of curvature of the surface. The same equation applies for a bubble of gas within a liquid, with the consequence that the vapor pressure p is larger for small bubbles according to the relation

$$\ln \frac{p}{p_0} = \frac{2\gamma}{r\rho} \frac{M}{RT}$$

where p_0 is the vapor pressure over a liquid surface of infinite radius, R is the gas constant, and M is the molecular weight. *See* VAPOR PRESSURE.

Detergents, soaps, and flotation agents owe their usefulness to their ability to lower the surface tension of water, thereby stabilizing the formation of small bubbles of air. At the same time, the interfacial tension between solid particles and the liquid phase is lowered, so that the particles are more readily wetted and floated after attachment to air bubbles. *See* FLOTATION; INTERFACE OF PHASES; PARACHOR; SURFACE-ACTIVE AGENT. [S.H.S.]

Bibliography: N. K. Adam, *The Physics and Chemistry of Surfaces*, 3d ed., 1941; S. Glasstone *Textbook of Physical Chemistry*, 2d ed., 1946

Surface water

A term commonly used to designate the water flowing in stream channels. The term is sometimes used in a broader sense as opposed to "subsurface water." In this sense, surface water includes water in lakes, marshes, glaciers, and reservoirs as well as that flowing in streams. In the broadest sense surface water is all the water on the surface of the earth and thus includes the water of the oceans Subsurface water includes water in the root zone of the soil and ground water flowing or stored in the rock mantle of the earth. Subsurface water differs from surface water in the mechanics of its movement as well as in its location.

Surface and subsurface water are two stages of the movement of the earth's water through the hydrologic cycle The world's ocean and atmospheric moisture are two other main stages of the grand water cycle of the earth. At any time there is a certain quantity of water in the various stages of the cycle. For example, the 11.5×10^9 acre-ft of water in the atmosphere is much greater than the 0.25×10^9 acre-ft of water in the stream channels *See* HYDROLOGY.

Considerably more may be outlined concerning water's rate of movement through the several stages of the hydrologic cycle In some stages, glaciers for example, water is locked up for long periods of time; but water in the atmosphere or in the streams is transient. A numerical value for the time of transit of water is its detention period in years—specifically, the ratio between the bulk or volume of water in a given stage of the hydrologic cycle and its mean rate of flow through that stage For example, the earth's supply of water as a whole amounts to about $1{,}100{,}000 \times 10^9$ acre-ft. The mean

rate of flow through the hydrologic cycle is of the order of 320 × 10⁹ acre-ft per year. Thus the detention period is about 3000 years; that is, on the average, each particle of the earth's supply of water partakes in movement through that cycle once in 3000 years. This is an average—some particles may move more than once, some part way, and some not at all in this period of time

The following table gives estimates of the amounts of water in various parts of the hydrologic cycle and their detention periods.

Distribution of the world's supply of water

Location	Volume of water 10^6 acre-ft	Percentage of total	Detention period years
World's oceans	1 060 000	97 39	5 000
Surface water on the continents			
Glaciers and polar ice caps	20 000	1 83	2 000
Fresh-water lakes	100	0 0095	100
Saline lakes and inland seas	68	0 0063	50
Average in stream channels	0 25	0 00002	0 05
Total surface water	20 200		100 av
Subsurface water on the continents			
Root zone of the soil	10	0 00093	0 25
Ground water above 2500 ft	3 700	0 339	5
Ground water below 2500 ft	4 600	0 425	100
Total subsurface water	8 300		
Atmospheric water	11 5	0 0011	0 03
Total world water (rounded)	1 088 000	100	3 000

It may be noted that surface water on the continents is but a small part of the world's water and that the bulk of that is in fresh-water lakes However, the detention period is also short This means that the surface-water part and especially the water in the streams, is rapidly discharged and replenished That is why surface water as well as the shallower ground water is called a renewable resource Water that has a detention period of more than a generation is not renewed within sufficient time to be so considered

Source of water in streams. Precipitation that reaches the earth is subdivided by processes of evaporation and infiltration into various routes of subsequent travel. Evaporation from wet land surfaces and from vegetation returns some of the water to the atmosphere immediately. Precipitation that falls at rates less than the local rate of infiltration enters the soil. Some of the infiltrated water is retained in the soil sustaining plant life and some reaches the ground water.

Because of the slope of the land surface, the precipitation that exceeds the capacity of the soil to absorb water flows overland in the direction of the steepest slope and concentrates in rills and minor channels. During storms most of the water in surface streams is derived from that portion of the precipitation which fails to infiltrate the soil. In some forested areas of high relief and probably in some other areas, stormflow in stream channels is composed, in large part, of water which was infiltrated into the surface soil but which moved rapidly through the surficial mantle of litter and humus to the channels

When streams are low, on the other hand, the bulk of the water in channels is the contribution of ground water derived from precipitation that infiltrated during storms. The flow of surface streams during rainless periods represents the gradual draining of water stored temporarily in the ground. Dry-weather streamflow is the overflow of a ground-water reservoir.

The distinction between surface and subsurface water, though useful, should not obscure the fact that water on the surface and water underground is physically connected through pores, cracks, and joints in rock and soil material. In many areas, particularly in humid regions, surface water in stream channels is the visible part of a *reservoir*, which is *partly underground*; the water surface of a river is the visible extension of the surface of the ground water

Disposition of precipitation. Streamflow represents only a small percentage, on the average, of the water that falls as precipitation The flow in streams under natural conditions is called runoff The ratio of average annual runoff to average annual precipitation in the United States ranges from 20 to 40% in humid parts of the United States and from 2 to 4% in semiarid areas. On the average, the annual budget of water over the United States is roughly as follows:

Average precipitation	30 in
Runoff by rivers to sea	9 in.
Evapotranspiration from plants and soil	21 in
Transport of atmospheric moisture from oceans to continental area	9 in.

The 30 in. of water contributed by precipitation must be balanced by a return of water to the atmosphere. There are 21 in returned to the atmosphere by evapotranspiration from the continental area and 9 in. flow to the oceans. Thus, to balance the atmospheric budget, the 9 in of water that is transported as vapor from the oceans to the continents must be included

Average runoff There are great geographic variations within this average balance as can be visualized by a map showing annual runoff in the United States The total runoff is greatest in areas of highest precipitation and lowest losse[illegible]

[illegible] part of the high mountain country

Much of the semiarid parts of New Mexico, Arizona, and parts of California and Nevada yield an annual runoff of only a few tenths of an inch from a precipitation of 4–8 in. In the mountainous parts of the same areas, precipitation reaches 25–30 in

lies between 90° and 180°. Three interfaces exist when a droplet of liquid contacts a solid, and three corresponding interfacial tensions exist: γ_{SL}, γ_{SV}, and γ_{LV}. The subscripts S, L, and V refer to solid, liquid, and vapor. At equilibrium, a balance of interfacial tensions exists at the line of common contact, which intersects the figures at point O. *For the case of a liquid which wets the solid* ($\theta < 90°$), this equilibrium is expressed by the relation:

$$\gamma_{SV} = \gamma_{SL} + \gamma_{LV} \cos \theta$$

Capillarity. Liquids which wet the walls of fine capillary tubes rise to a height which depends upon the tube radius, the surface tension, the liquid density, and the contact angle. In Fig. 2, a liquid of density ρ is shown as having risen to a height h in a capillary whose radius is r. A balance exists between *the force exerted by gravity on the mass of* liquid raised in the capillary and the opposing force caused by surface tension. The former is $\pi r^2 h \rho g$, whereas the latter is $2\pi r \gamma$, assuming the contact angle *to be zero. It is clear that* $h = 2\gamma / r\rho g$, and that the capillary rise varies inversely with the tube radius and the liquid density. Liquids which do not wet the capillary walls are depressed in height according to the same equation.

The shape of the free surface of a liquid in a vessel is only an approximation to a plane. In narrow tubes the meniscus of a liquid is concave upward if the liquid wets the tube, and conversely convex upward if it does not wet the tube. A *pressure difference* exists between the concave and convex sides of the surface, the excess pressure on the concave side over the convex side being given by the relation

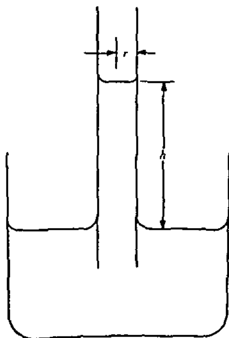


Fig. 2. Rise of liquid in capillary tube.

$$p = \frac{1}{r_1} + \frac{1}{r_2}$$

where r_1 and r_2 are the principal radii of curvature of the surface. The same equation applies for a bubble of gas within a liquid, with the consequence that the vapor pressure p is larger for small bubbles according to the relation

$$\ln \frac{p}{p_0} = \frac{2\gamma}{r\rho} \frac{M}{RT}$$

where p_0 is the vapor pressure over a liquid surface of infinite radius, R is the gas constant, and M is the molecular weight *See* Vapor pressure.

Detergents, soaps, and flotation agents owe their usefulness to their ability to lower the surface tension of water, thereby stabilizing the formation of small bubbles of air. At the same time, the inter*facial tension between* solid particles and the liquid phase is lowered, so that the particles are more readily wetted and floated after attachment to air bubbles *See* Flotation; Interface of phases; Parachor; Surface-active agent. [V.B.Y.]

Bibliography: N. K. Adam, *The Physics and Chemistry of Surfaces*, 3d ed., 1941; S. Glasstone, *Textbook of Physical Chemistry*, 2d ed., 1946.

Surface water

A term commonly used to designate the water flowing in stream channels. The term is sometimes used in a broader sense as opposed to "subsurface water." In this sense, surface water includes water in lakes, marshes, glaciers, and reservoirs as well *as that flowing in streams.* In the broadest sense, surface water is all the water on the surface of the earth and thus includes the water of the oceans. Subsurface water includes water in the root zone of *the soil and ground water flowing or* stored in the rock mantle of the earth. Subsurface water differs from surface water in the mechanics of its movement as well as in its location.

Surface and subsurface water *are two* stages of the movement of the earth's water through the hydrologic cycle. The world's ocean and atmospheric moisture are two other main stages of the grand water cycle of the earth. At any *time there* is a certain quantity of water in the various stages of the cycle. For example, the 11.5×10^9 acre ft of water in the atmosphere is much greater than the 0.25×10^9 acre-ft of water in the stream channels *See* Hydrology

Considerably more may be outlined concerning water's rate of movement through the several stages of the hydrologic cycle. In some stages, glaciers for example, water is locked up for long periods of time, but water in the atmosphere or in the streams is transient. A numerical value for the time of transit of water is its detention period in years—specifically, the ratio between the bulk or volume of water in a given stage of the hydrologic cycle and its *mean rate of flow* through that stage For example, the earth's supply of water as a whole amounts to about $1,100,000 \times 10^9$ acre-ft. The mean

rate of flow through the hydrologic cycle is of the order of 320 × 10[9] acre-ft per year. Thus the detention period is about 3000 years; that is, on the average, each particle of the earth's supply of water partakes in movement through that cycle once in 3000 years. This is an average—some particles may move more than once, some part way, and some not at all in this period of time.

The following table gives estimates of the amounts of water in various parts of the hydrologic cycle and their detention periods.

Distribution of the world's supply of water

Location	Volume of water 10[6] acre-ft	Percentage of total	Detention period, years
World's oceans	1 060 000	97.39	5 000
Surface water on the continents			
Glaciers and polar ice caps	20 000	1.83	2 000
Fresh-water lakes	100	0.0093	100
Saline lakes and inland seas	68	0.0063	50
Average in stream channels	0.25	0.00002	0.03
Total surface water	20 200		700 av
Subsurface water on the continents			
Root zone of the soil	10	0.00091	0.25
Ground water above 2500 ft	3 700	0.339	5
Ground water below 2500 ft	4 600	0.423	100
Total subsurface water	8 300		
Atmospheric water	11.5	0.0011	0.03
Total world water (rounded)	1 088 000	100	3 000

It may be noted that surface water on the continents is but a small part of the world's water and that the bulk of that is in fresh-water lakes. However, the detention period is also short. This means that the surface-water part, and especially the water in the streams, is rapidly discharged and replenished. That is why surface water, as well as the shallower ground water, is called a renewable resource. Water that has a detention period of more than a [illegible]

[illegible]

re[illegible] is subdivided by processes of evaporation and infiltration into various [illegible] of surface [illegible]

[illegible] precipitation that falls at rates less than the local rate of infiltration enters the soil. Some of the infiltrated water is retained in the soil, sustaining plant life, and some reaches the ground water.

Because of the slope of the land surface, the precipitation that exceeds the capacity of the soil to absorb water flows overland in the direction of the steepest slope and concentrates in rills and minor channels. During storms most of the water in surface streams is derived from that portion of the precipitation which fails to infiltrate the soil. In some forested areas of high relief and probably in some other areas, stormflow in stream channels is composed, in large part, of water which was infiltrated into the surface soil but which moved rapidly through the surficial mantle of litter and humus to the channels.

When streams are low, on the other hand, the bulk of the water in channels is the contribution of ground water derived from precipitation that infiltrated during storms. The flow of surface streams during rainless periods represents the gradual draining of water stored temporarily in the ground. Dry-weather streamflow is the overflow of a ground-water reservoir.

The distinction between surface and subsurface water, though useful, should not obscure the fact that water on the surface and water underground is physically connected through pores, cracks, and joints in rock and soil material. In many areas, particularly in humid regions, surface water in stream channels is the visible part of a reservoir, which is partly underground; the water surface of a river is the visible extension of the surface of the ground water.

Disposition of precipitation. Streamflow represents only a small percentage, on the average, of the water that falls as *precipitation*. The flow in streams under natural conditions is called runoff. The ratio of average annual runoff to average annual precipitation in the United States ranges from 20 to 40% in humid parts of the United States and from 2 to 4% in semiarid areas. On the average, the annual budget of water over the United States is roughly as follows:

Average precipitation	30 in.
Runoff by rivers to sea	9 in.
Evapotranspiration from plants and soil	21 in.
Transport of atmospheric moisture from oceans to continental area	9 in.

The 30 in. of water contributed by *precipitation* must be balanced by a return of water to the atmosphere. There are 21 in. returned to the atmosphere by evapotranspiration from the continental area and 9 in. flow to the oceans. Thus, to balance the atmospheric budget, the 9 in. of water that is transported as vapor from the oceans to the continents must be included.

Average runoff. There are great geographic variations within this average balance as can be visualized by a map showing annual runoff in the United States. The total runoff is greatest in areas of highest precipitation and lowest losses. In the mountains of the Northwest, where precipitation is as much as 150 in. annually, the runoff in surface streams is more than 50 in. annually over a considerable part of the high mountain country.

Much of the semiarid parts of New Mexico, Arizona, and parts of California and Nevada yield an annual runoff of only a few tenths of an inch from a precipitation of 4–8 in. In the mountainous parts of the same areas, precipitation reaches 25–30 in.

and, locally, the runoff may average as much as 10 in. annually from small areas Much of the United States west of the 100th meridian has an average annual runoff of less than 3 in.

Discharge in nearly all streams has a marked annual cycle. Spring and early summer are usually periods of high flow resulting from snowmelt and rain during a period of relatively low water-loss by transpiration. Late summer is often the period of lowest flow owing to the infrequency of precipitation and to the maximum use of water by leafy vegetation. The distribution of runoff throughout the year is not the same for the whole country, however, because it varies regionally depending on the seasonal distribution of precipitation and on the importance of snowmelt as a source of runoff.

Individual streams and rivers in the United States have been measured over varying periods of time at about 12,000 sites (*see* STREAM GAGING). Daily discharge at most of them is published by the U.S. Geological Survey (USGS) in the series of Water-Supply Papers, entitled *Surface Water Supply of the United States* The data are tabulated for each measuring station and are grouped in volumes by river basins.

If a stream goes dry occasionally or does not have enough flow to satisfy a desired use, storage reservoirs can be built to conserve high flows for release during low-flow periods The design of such reservoirs requires a knowledge of how low a flow is likely to be experienced, how long it may last, and how frequently it can be expected to recur Streamflow records can be analyzed to answer these questions.

Extremes of runoff. The amount of streamflow available during periods of extremely low flow can be shown by flow-duration curves and other types of low-flow frequency analysis. Flow-duration curves, which express the per cent of time the flow of a particular stream has been equal to or greater than any given quantity, have long been used in water-power studies, and curves showing the frequency and severity of annual lows are now being used in water-supply studies and stream-sanitation studies. The preparation of such curves is facilitated by the use of electronic computers Many such computations have been made and published, but there is no single source or systematic publication of such material. Summaries of published data on low flows for some rivers in the United States may be obtained from the USGS.

Much more information is organized and published on floods than on low flows Representative values of peak discharges showing the magnitude of extreme flows experienced in drainage basins of various sizes are given in the following table

Data similar to those included in the table are published in a systematic manner under the headings of "extremes" in the tabulated flow data for individual gaging stations in the Water-Supply Papers of the USGS Flood expectancy, even for ungaged places, may be estimated from curves the USGS is publishing in a series of papers dealing with the frequency and magnitude of floods by in[illegible] dividual states or areas. By 1959 these analyse[illegible] were complete for about half of the nation

Representative values of extreme peak flows

River	Date	Drainage area, mi²	Peak discharge cfs	Peak discharge cfs/mi²
Big Branch near Waynesville, N C	Aug 30, 1940	0 4	4 500	[illegible]
Big Creek near Waynesville, N C	Aug 30, 1940	1 32	12 900	[illegible]
Laurel Creek above White Pine, W. Va	August, 1943	2 42	7 400	[illegible]
Cameron Creek near Tehachapi Calif	Sept. 30, 1932	3 59	13 500	[illegible]
Unnamed Creek near York, Nebr	July 9, 1950	6 93	23 000	[illegible]
Meyers Canyon near Mitchell, Oreg	July 13, 1956	12 7	54 500	[illegible]
Alazan Creek, below Martinez Creek, Tex	Sept 9, 1921	17 1	25 900	[illegible]
Salem Creek below Woodstown, N J	Sept 1, 1940	17 5	26 100	[illegible]
Morgan Creek near Chapel Hill, N C	Aug 4, 1924	29 1	30 000	[illegible]
Pine Tree Canyon, 12 miles north of Mohave, Calif	Aug 12, 1931	35 0	59 500	[illegible]
Elkhorn Creek, Keystone, W Va	June, 1901	44	60 000	[illegible]
Little Nemaha River at Syracuse, Nebr	May 9, 1950	218	225 000	[illegible]
Guadalupe River near Ingram, Tex	July 1, 1932	336	206 000	613
W Nueces River near Brackettville, Tex	June 14, 1935	402	580 000	[illegible]
W Nueces River near Cline, Tex	June 14 1935	880	536 000	[illegible]
Eel River at Scotia, Calif	Dec 22, 1955	3,113	541 000	[illegible]
Devils River near Del Rio, Tex	Sept 1, 1932	4,060	597 000	[illegible]
Neosho River near Parsons, Kans	July 14, 1951	4 817	410 000	[illegible]
Little River at Cameron Tex	Sept 10, 1921	7,000	647 000	[illegible]
Ohio River at Sewickley Pa	Mar 18, 1936	19 500	574 000	[illegible]
Susquehanna River at Marietta Pa	Mar 19 1936	25,900	787,000	[illegible]
Ohio River at Evansville Ky	Jan 29 1937	107 000	1 410 000	[illegible]
Ohio River at Metropolis Ill	Feb 1, 1937	203,000	1 850 000	[illegible]
Columbia River at The Dalles, Oreg	June 6 1894	237 000	1,210 000	[illegible]

Relation of runoff to drainage area Averag[illegible] water yield or annual runoff in a physically homo[illegible] geneous area increases in direct proportion to th[illegible] size of the drainage basin, but this is not true o[illegible] flood potentiality Small drainage basins produc[illegible] larger peak flows per unit of drainage area than d[illegible] large basins. The relation between magnitude o[illegible] flood peak and the contributing drainage area ma[illegible] be expressed by the following equation:

$$Q = aA^c$$

where Q is discharge in cubic feet per second (cfs[illegible] A is drainage area in square miles (mi²), a and [illegible] are coefficients If Q represents the average annu[illegible] water yield then the value of c is approximatel[illegible] unity for most basins in humid areas. If Q repre[illegible] sents peak flood discharge of a given frequency o[illegible] recurrence interval, such as a 10-year or 50-yea[illegible]

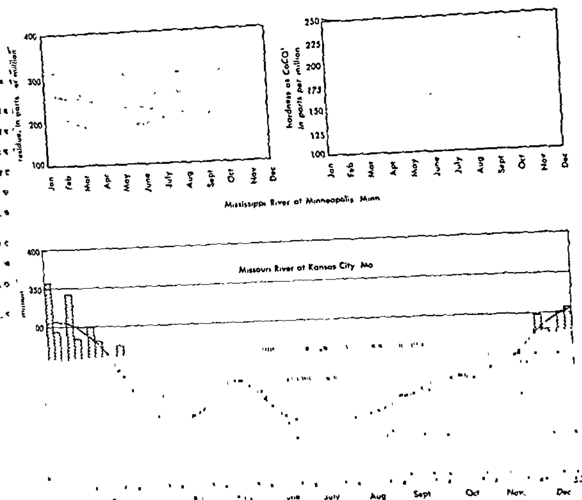


Fig 1 Seasonal variation in quality characteristics of streams.

average recurrence interval, then past observations indicate that the value of c is between 0.7 and 0.8 for a large variety of drainage basins

Some reasons for these differences in exponents are as follows Each square mile of a homogeneous drainage basin contributes an *equal quantity* of water over the course of many years *Therefore*, the average annual runoff from a drainage basin is simply the summation of the contribution of each unit area

Peak rates of runoff during floods, on the other hand, may be viewed roughly as the contributions from the parts of the drainage basin at varying distances away The different distances mean that the peak flood contributions reach the several points downstream at *different times*, and the farther downstream one goes, the greater the flood wave is spread out The flattening action is increased by the temporary *storage* of water in the stream channels and in the bordering flood plains As it moves downstream, the flood crest is increased by the contributions of tributaries, but because these contributions are not synchronized, the peak contributions are not *simply additive* Owing to the delays due to channel storage, flood peak discharges increase with drainage area, but to a power less than unity; peak flows per unit of drainage area decrease with drainage area

Quality of water. The usefulness of available water is often limited by its quality. Good quality often is considerably more important than unlimited quantity, particularly in industry

It is characteristic of river waters to vary in chemical and physical quality almost continuously. The chemical quality of the water in lakes, particularly large ones remains relatively constant throughout the year Differences between streams are caused by several factors These include (1) the nature of soils and rocks over and through which the water flows, (2) the length of time the water is in contact with various rock types, (3) the water quality of tributary streams, (4) proportion of flow due to ground-water discharge (5) flood [illegible]

Figure 2 illustrates variation in dissolved solid[illegible] with streamflow. In general, streams flowing int[illegible] the Atlantic Ocean, eastern Gulf of Mexico, and th[illegible] North Pacific Ocean are of good to excellent qua[illegible]ity for general purposes. Dissolved matter an[illegible]

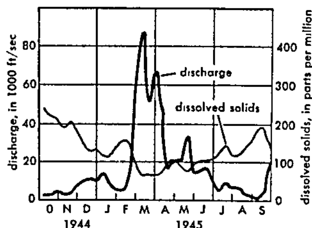


Fig. 2. Discharge and dissolved solids in Allegheny River at Kittanning, Pa.

hardness are usually below 100 parts per million (milligrams of dissolved substance in a liter of water) and often below 50. This applies to streams in their natural state and does not take into account the effects of pollution.

Midcontinent and southwestern streams generally have high concentrations of dissolved matter, some excessively so, and must be treated extensively for a large variety of general uses. For example, the water of the Missouri River at Kansas City has about 220–500 ppm of dissolved matter and hardness of about 190–300 ppm during a typical year. A generalized picture of the variations in surface-water quality throughout the United States is given in Fig. 3.

The temperature of streams varies continuously throughout the year, ranging from a minimum of freezing (about 32°F) in winter in northern latitudes to a maximum of about 90°F in summer in southern latitudes. The monthly average water temperature generally follows rather closely the monthly average air temperatures except in areas where the flow is made up largely of melting snow or ice, or of ground water. The water temperature of streams fed by snowmelt is lower than the air temperature for some distances downstream from the snow fields. The water temperature of streams whose flow comes from ground water tends to be more uniform and in summer months is usually colder than the average air temperature

Sediment also affects surface-water quality. Nearly all streams are turbid during flood periods, some carrying tremendous quantities of sediment which must be removed before the water is suitable for industrial and most other uses In eastern United States the amount of suspended matter in typical streams seldom exceeds 0.3% (3000 ppm) and generally averages a few hundred parts per million. In many midcontinent and western streams sediment concentrations are much greater, a maximum of 10% not being uncommon in some streams; frequently the maximum is considerably higher.

The sediment-carrying characteristics of western streams range from relatively clear-flowing mountain streams to near mud flows in the intermittent streams of arid regions. The sediment concentration of the Colorado River at Grand Canyon, Ariz., averaged about 0.6% (6000 ppm) for a period of nearly 30 years.

Data on the chemical and sediment loads of streams in the United States are published systematically in the Water-Supply Papers of the USGS.

Characteristics of river channels. Rivers and streams form channels which are the routes of transport for water and debris-load delivered to them by the basin during the slow process of landscape degradation. Channels have certain characteristics that are amazingly universal regardless of the location of the river basin. The basic mechanisms which lead to these common characteristics are only imperfectly understood.

Water only partly fills the channel during periods of low flow. Generally, on more than half the days in a year, only the lowest one-fifth of the channel depth is filled with water. The channel flows bankful about once a year on the average. Though this varies somewhat from reach to reach and from one river basin to another, the generalization that the channel is constructed by the river so that it overflows once every year or every two years is one of the most interesting and potentially useful items of information about natural channels Because a flood is, by definition, a flow which exceeds channel capacity, the above generalization emphasizes the fact that flooding is a natural characteristic of rivers.

Natural channels tend to be roughly trapezoidal rather than elliptical or semicircular. The width of the water surface along the channel at any given frequency of flow (high flow or low flow) generally

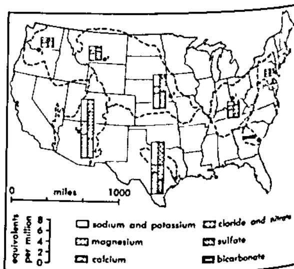


Fig 3 Chemical composition of water in several basins Data are for the following: (1) Deleware River at Trenton, N.J., 1951; (2) Savannah River near Clyo, Ga., 1940; (3) Ohio River at Cincinnati, Ohio, 1951, (4) Brazos River at Richmond, Tex., 1951; (5) Colorado River near Grand Canyon, Ariz., 1951; (6) Columbia River near Rufus, Oreg., 1951; (7) Yellowstone River at Billings, Mont., 1951; and (8) Missouri River at Nebraska City, Neb., 1951.

increases as the square root of the discharge, as discharge increases downstream with the addition of tributaries The shape of the channel cross section is asymmetric at bends, the deepest part being the concave bank

e depth increases downstream but not as rapas the width The width-depth ratio increases stream as about the 0.1 power of the disge, or

$$\frac{w}{d} \propto Q^{1}$$

re w and d are respectively mean width and n depth, and Q is discharge of a given freacy

hannel slope or gradient decreases downstream erally following an exponential or logarithmic Also, size of debris composing the bed tends diminish downstream. Despite the decreasing pe in the downstream direction, mean water ocity increases slightly along the length of a er or maintains a roughly constant value

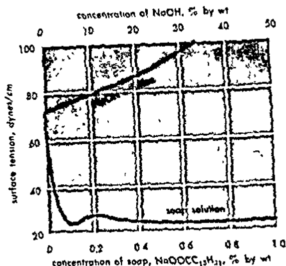
(L.B.L.)

Bibliography: W. G Hoyt and W B. Langbein, oods, 1955, W B Langbein, *Annual Runoff in e United States*, USGS Circ 52, 1949; L. B Leold and W. B Langbein, *A Primer on Water*, SGS, 1959, L. B Leopold and T Maddock, Jr, *he Hydraulic Geometry of Stream Channels and me Physiographic Implications*, USGS Profesaper 252, 1953, O. E Meinzer (ed), *Hydrology*, 349, R L. Nace, *Water Management, Agriculture, nd Ground-Water Supplies*, Annual Meeting, Am ssoc Advance. Sci, 1958

Surface-active agent

A substance that even though present in small amounts, exerts a marked effect on the surface behavior of a system These agents are essentially responsible for producing great changes in the surface energy of liquid or solid surfaces, and their ability to cause these changes is associated with their tendency to migrate to the interface between two phases Consequently surface-active agents are of potential interest wherever there are solid solid solid-liquid solid gas, liquid-liquid or liquid gas interfaces, and of particular interest at liquid gas interfaces at which the surface-active agent is a solute whose presence makes the surface properties of the solution greatly different from those of the solvent See INTERFACES OF PHASES

Mechanism. Soap for example, when dissolved in small quantities in water, is responsible for greatly decreasing the surface tension of water, and it is this property of soap that accounts for its ability to act as a detergent See SOAP AND DETERGENT In contrast to soap and other related substances that lower the surface energy of a liquid other solutes, such as inorganic salts, acids, and bases may increase the surface tension of a liquid (see illustration), but their effect in increasing the surface tension is not nearly so great as the effect of those agents that decrease the surface tension



Effect of surface-active agents on the surface tension of water.

Occasionally the term surface-inactive solutes is applied to these substances whose presence causes an increase in surface tension.

The importance of surface-active agents is indicated by their strategic necessity in such processes as lubrication, wetting, foaming, emulsification, detergency, water repellence, water proofing, spreading, and dispersion. In lubrication, for example, the oiliness of a hydrocarbon oil can be improved by the addition of a surface-active agent. In order to achieve lubrication between two solid surfaces, a thin film of liquid must be preserved in the space between the two solid surfaces. The viscosity of this liquid film and the ability of the liquid to wet the solid surfaces determine the resistance of the lubricant system to being squeezed mechanically from the region between the two solid surfaces. Addition of fatty acids, fatty oils, metallic soaps, and various derivatives of aromatic and aliphatic hydrocarbons commonly improves the lubricant qualities of mineral hydrocarbons, and these additives are truly surface-active agents

The mechanism by which surface-active agents alter the surface energy of a solid or liquid is attributed to the dual nature of the molecules or ions of these substances Within a single molecule or ion of a surface-active agent, there is a group that is lyophilic toward the dispersing medium or solvent, and at a suitable distance within the same molecule or ion, there is another group that is lyophobic toward the dispersing medium This ability to embody within the same molecular particle two different groups whose properties are diametrically opposed is sometimes termed amphipathy. For example, the surface activity of sodium oleate, $NaOOCC_{15}H_{31}$, is attributed to the combined effect of the hydrophilic ionic carboxyl salt group at one end of the molecule and the hydrophobic hydrocarbon group that constitutes the remainder of the molecule In a dilute solution of sodium oleate, the solute migrates to the surface

where the hydrophobic parts of the molecules can achieve their lowest energy positions as the result of the solvent's striving to exclude the hydrocarbon group from the solution. Even though the external phase is gaseous, the hydrophobic groups find a sufficiently sympathetic environment at the surface of the liquid. If, on the other hand, the external phase were an oil, the hydrocarbon groups would find an even more sympathetic environment, and in either event, the surface energy of the original solvent would be greatly diminished.

Classification. Surface-active agents are usually classified in three groups, anionic, cationic, and nonionic types. Anionic types include carboxylate ions such as occur in sodium oleate. The carboxyl group may be attached directly to the hydrophobic group, or there may be an intermediate ester, amide, or sulfonamide linkage. There are a large number of anionic agents derived from sulfuric and sulfonic acids in which the hydrophobic groups attached to them include aliphatic and aromatic groups that often contain substituents of varying polarity, such as halide, hydroxyl, ether, and ester groups.

Cationic surface-active agents are usually derived from the amino group where, through either primary, secondary, or tertiary amine salts, the hydrophilic character may be achieved by aliphatic and aromatic groups that may be altered by substituents of varying polarity. Other nitrogen compounds, such as quaternary ammonium compounds, guanidine, and thiuronium salts, are included in the cationic class.

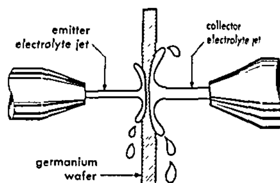
The third class of surface-active agents, the nonionic type, are organic substances which contain groups of varying polarity and which render parts of the molecule lyophilic, whereas other parts of the molecule are lyophobic. Examples include polyethylene glycol, polyvinyl alcohol, polyethers, polyesters, and polyhalides. In this class are often included certain colloidal substances such as graphite, powdered metals, metallic oxides, clays, macromolecules, and polymers. *See* FLOTATION; LUBRICANT; SURFACE TENSION. [W H S]

Bibliography: J. L. Moillet and B. Collie, *Surface Activity*, 1951; A. M. Schwartz and J. W. Perry, *Surface Active Agents*, 1949.

Surface-barrier transistor

A transistor in which the emitter and collector are formed on opposite sides of a semiconductor wafer by a combination of jet etching and plating techniques. The material is usually *n*-type germanium. Improved performance results from slight pulse heating of the plated contacts, so that in practice there may be microalloyed or diffused emitter and collector regions under the plated contacts, giving a *p-n-p* type transistor structure. These transistors are high-frequency units which range up to 150 megacycles. They have a low power-dissipation rating of about 30 milliwatts.

The method of fabrication is unique and lends itself to automation. A wafer of *n*-type germanium is



Surface-barrier jet etching and plating technique

mounted between two very small nozzles which train two jets of electrolyte against its opposite surfaces. The wafer is electrically biased positive with respect to the electrolyte. As the etching proceeds and the web of germanium becomes thin the resistance to flow of etching current increases. At a predetermined value the polarity of the germanium is reversed and appropriate metal ions in the electrolyte are plated out on the freshly etched surface of the wafer. This method gives good control of thin base regions and therefore lends itself well to minimizing the base transit time of injected carriers.

The abrupt nature of the plated barriers tends to give a high collector capacitance, and most units show a lower current gain than a normal *p-n-p* transistor. *See* TRANSISTOR. [L.P.H.]

Bibliography: J. W. Tiley and R. A. Williams, Electrochemical techniques for fabrication of surface-barrier transistors, *Proc. IRE*, 41:1706-1708, 1953.

Surgery

The field of medicine concerned chiefly with the treatment of disease or injury by means of operative or manipulative procedures.

Crude stone instruments were used by primitive man for trephining, blood-letting, and other procedures. These were gradually replaced by bronze, then by iron and steel instruments as these metals appeared. Various civilizations, such as the Egyptian, Babylonian, Hindu, and Chinese, advanced to certain levels of surgical proficiency at different times. Hippocrates separated science from religion and philosophy, innovated observation and reason, and compiled the ethical code still used by physicians. Galen, although a skilled practitioner, left scientific method to write voluminously in the questionable light of pragmatism.

For the next 1700 years Galen's authority was final in all things medical and surgical. The Dark Ages of medieval times suppressed the new and the questioning, although certain innovations were encouraged if Christian motives were involved. Great universities, hospital systems, and medical legislation arose and the development of medicine and surgery proceeded sporadically. Surgery became divorced from medicine and was often relegated to barbers and [illegible]ers because the

learned did not wish to handle the sick and the maimed. Here and there an individual surgeon made his often obscure contribution.

From the Renaissance into the nineteenth century great figures in surgery arose and new techniques and principles were developed. The relief of pain through use of anesthesia, introduced in 1846, initiated the modern period of technical advancement

Anesthesia was soon followed by Robert Lister's antiseptic techniques in 1867, based partially on Louis Pasteur's work with microbes During the great wars of the nineteenth and twentieth centuries, surgical knowledge advanced by leaps and bounds

Since 1930 the prevention and treatment of [illegible]

The increas[illegible]

gical specialty are required after the completion [illegible]

[illegible]ently recognized surgical specialties include general surgery, thoracic, plastic orthopedic, and neurosurgery in addition to the fields of opthalmology, otolaryngology, proctology, and urology. In each of these surgery plays a prominent role but other forms of treatment are of course also employed Despite this high degree of specialization, there is close correlation between the surgeon and other medical specialists and general practitioners. Adjunct specialties such as anesthesiology and physical medicine at the physician level, have developed, in addition to large groups of specially trained technical and nursing personnel *See* ANESTHETIC

The general practitioner often skilled in minor or major surgery is still the mainstay of most smaller communities There is a current trend to encourage the development of the well trained general physician rather than over-specialization for every medical student.

Advances in surgery since 1946 have had many dramatic and far-reaching implications Corrective surgery of the blood vessels and of the heart, including the use of stored tissue or prosthetic devices, is aiding patients who formerly would have died or remained chronically ill Plastic surgery has allowed the rehabilitation of thousands maimed by war, disaster and accidents The use of surgery in the treatment of malignancy remains the only treatment so far in which definitive results can be seen Neurosurgery is developing procedures for the treatment of nervous diseases formerly thought to be hopeless Basic investigative work has brought about the methods for preservation and storage of blood bone, and other tissues to be used when needed *See* PROSTHESIS. [E.C.ST.]

Surging

A sudden and momentary change of voltage or current in a circuit. It can be due to a sudden change in the applied input signal, a sudden change in the load placed on the circuit, or to the action of a relay, switch, or other device that changes operating conditions within the circuit. The resulting surges or transients are often called pulses or impulses when they have only one polarity An oscillatory surge includes both positive and negative polarity values. Surging in electric circuits corresponds to overshooting. Cathode-ray oscilloscopes are frequently used to obtain visual patterns of the transient voltages due to surging. *See* TRANSIENT, ELECTRIC. [J.M.R.]

Surveillance radar

Ground radar used for traffic control purposes in the approach and landing zone. This radar is used primarily to assist controllers in converting random arrivals to regular landings and in positioning such aircraft so that they may make low approaches by the use of a fixed beam, low-approach system or a precision radar low-approach system. *See* AIR TRAFFIC CONTROL. [P.C.S.]

Surveying

The measurement of dimensional relationships among points, lines, and physical features on or near the earth's surface.

Basically, surveying determines horizontal distances, elevation differences, directions, and angles These basic determinations are applied further to computation of areas and volumes and to establishment of locations

Surveying is typically applied to such jobs as locating and measuring property lines; laying out a new dam, building, road, or pipe line; and mapping the topography of an area. *See* CIVIL ENGINEERING

Horizontal measurements usually are assumed to be parallel to a common plane. Each measurement



Fig 1 Subtense bar. (Lockwood, Kessler, and Bartlett, Inc)

has both length and direction. Length is expressed in feet or in meters. Direction is expressed as a bearing of azimuthal angle relationship to a reference direction (*see* AZIMUTH). The degree-minute-second system of angular expression is standard in the United States.

Reference, or control, is a concept that applies to the positions of lines as well as to their directions. In its simplest form, the position control is an identifiable point of origin for the line or lines of a survey. Broader utility derives from a system of coordinates, wherein each point's position is expressed in terms of a distance north and a distance east of a point of origin located south and west of the survey area.

Coordinate systems may be assumed, or established coordinate systems may be used. The most widely adaptable systems in the United States are the State Plane Coordinate systems, based on planes established by the U.S. Coast and Geodetic Survey (USCGS).

Vertical measurement adds the third dimension to an object's position. This dimension is expressed as the distance above some reference surface, usually mean sea level, called a datum Mean sea level is determined by averaging high and low tides during at least a lunar month (*see* SEA LEVEL FLUCTUATIONS).

Precision. Surveying devices and procedures possess individual limitations for accuracy of measurement. The term precision expresses the idea of *degree of accuracy*

The choice of methods depends on the precision required for the result. Surveys of valuable property or major bridge layouts demand higher orders of precision than, for example, a preliminary highway profile.

Instrumental surveys are designated first-order, second-order, third-order, and fourth-order. First-order is maximum precision; fourth is minimum. Established criteria differentiate the orders of precision.

Horizontal control. The main framework, or control, of a survey is laid out by two main techniques, traverse and triangulation. Electronic trilateration also is used.

In traverse, adopted for most ordinary surveying, a line or series of lines is established by directly measuring lengths and angles In triangulation, used mainly for large areas, angles are again directly measured, but distances are computed trigonometrically. This necessitates triangular patterns of lines and starting from a base line of known length. New base lines are measured at intervals Trigonometric methods are also used in electronic trilateration, but lengths, rather than angles, are measured by electronic range-finding instruments

Traversing. Traverse lines are usually laid out as a closed polygon. This makes it possible to check the accuracy of the field work by computing how nearly the figure closes mathematically. Commonly there are both angular and distance errors of closure. Angular error is checked by geometric theory The sum of all interior angles in any polygon must equal the number of angles, minus two, times 180. The distance error may be determined by resolving the field measurement for each line into its north-south and east-west components, called latitudes and departures, respectively. For the figure to close, components bearing north must equal in total length those bearing south, and components bearing east must equal those bearing west Errors calculated along the north-south and east west axes are the two components of the linear error of closure. If it is within the probable error limit for the order of precision chosen for the survey, the total error is apportioned to the survey lines.

Triangulation. This begins with the selection of points whose connecting sight lines form one triangle or a series of triangles. In a series, each triangle has at least one side common with each adjacent triangle. An initial side length, the base line, is measured, as are all angles. By successive application of the sine law, all remaining side lengths are computed Side directions also are carried forward, so that the relative positions of all points can be determined.

Trilateration. In electronic trilateration the figures are selected as for triangulation. An initial direction is determined and all side lengths are measured Interior angles are computed by oblique triangle formulas to obtain geometric checks on distances and to establish triangle-side directions Ensuing latitude and departure computations yield the relative positions of angle points.

Distance measurement. Traverse distances are usually measured with a tape but are also measured by stadia, subtense, trig-traverse, and electronic methods

Taping. The surveyor's tape, a steel ribbon or wire, is usually 100 ft long. However, it is available in other footages and in multiples of the 66-ft chain

Whether on sloping or level ground, it is horizontal distances that must be measured Horizontal components of hillside distances are measured by raising the downhill end of the tape to the level of the uphill end. On steep ground this technique is used with shorter sections of the tape. The raised end is positioned over the ground point with the aid of a plumb bob.

Where slope distances are taped along the ground, the slope angle can be measured with the hand-held clinometer (*see* CLINOMETER). The horizontal distance can then be computed or read from a table Precision in tape measuring is increased by refinements, such as standardization, *sag* correction, tension control, and thermal-expansion correction

Stadia The stadia technique requires no tape. A graduated stadia rod is held upright on a point and sighted through a transit telescope set up over another point *See* TRANSIT (ENGINEERING). The distance between the two points is determined from the length of rod visible between two horizontal hairs in the telescope *See*



2 Microwave distance measurement (Tellurom- Inc)

ubtense. In the subtense technique a horizontal of fixed length with targets at each end is set at the forward point (Fig 1). The transit, oc- ying the rear point, measures the horizontal an- subtended by the targets. Computation of the celes triangle's altitude yields the distance more isely than stadia up to about 500 ft, but it loses ision at greater distances because of the de- sing effect that distance changes have on the tended angle

rig traverse. In trig-traverse the subtense bar is laced by a measured base line extending at a t angle from the survey line; the base line is enough relative to the distance to assure the r of precision desired

lectronic methods Medium- to long-distance surements may be speeded by use of modu- d infrared light or radio microwaves The time gnal takes to travel to a distant receiver and rn to the sending instrument is measured and verted to distance Relative accuracy improves distance, because errors are largely constant

ngular measurement. The most common in- ment for measuring angles is the transit. It is ntially a telescope that can be rotated a meas- d angular amount about a vertical axis and a izontal axis Graduated circular plates concen- with each of the axes are the means of measure- t.

he transit is centered over a point with the aid ither a plumb bob suspended from the vertical or (on some transits) an optical plummet, ch enables the transitman to sight along the nsit's vertical axis to the ground through auxili- lenses

Horizontal angles. To measure a horizontal an- gle between two intersecting lines, the transit is set up over the intersection. The telescope is sighted along one of the lines, the graduated horizontal cir- cle is clamped against rotation, and the telescope is rotated to sight along the other line. The angle is indicated on the horizontal circle by another con- centric circular plate, inscribed with verniers, that rotates with the telescope The angle can be read directly if the initial reading has been set at zero Otherwise, the angle is computed as the difference between the initial and final readings.

To lay off a predetermined angle from some ref- erence line, the initial sight is taken along the line, the telescope is rotated through the angle desired, and a stake or other marker is set on the new line of sight.

The special case of laying off a 180° angle is simply the extension of a straight line. It is done by backsighting along the reference line and rotat- ing the telescope on its horizontal axis (transiting) for sighting ahead

Vertical angles These are measured by rotating the telescope on its horizontal axis and reading the angle on the vertical circle and verniers. Vertical angles are usually measured from a horizontal ref- erence line.

Angular precision. Transit precision is denoted by the least count of the verniers. A 1-minute transit, for example, has a circle graduated to half-degrees, with the verniers measuring a thirtieth of the gradu- ation, or 1 minute.

Transits of United States manufacture rarely have least counts less than a half minute Certain foreign transits, embodying optical micrometers and the magnification of fine graduations etched on glass, have least counts of 1 second and less.

Elevation differences. Elevations may be meas- ured in conjunction with rod

urer

elev

D ... leveling. Most third-order and all second- and first-order measurements are made by differential leveling, wherein a horizontal line of sight of known elevation is intercepted by a gradu- ated standard, or rod, held vertically on the point being checked The transit telescope may establish the sight line, or a specialized leveling instrument may be used *See* LEVEL (SURVEYING). For ap- proximations a hand level may be used.

In differential le

... called the backsight or plus sight, is added to the bench mark elevation to establish the height of the instrument With the level remaining where it is, the rod then is moved to a forward point (turning point). The reading to that point, called the foresight or minus sight, is subtracted from the height of the instrument to yield the ground elevation of the turning point

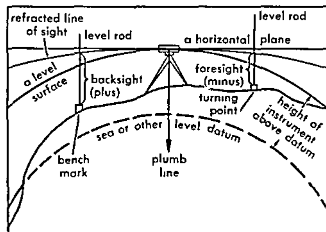


Fig. 3. Differential leveling theory

and effects of earth curvature and refraction of the line of sight. These are illustrated in exaggerated form in Fig 3.

The foregoing process is duplicated successively as necessary to obtain the desired new elevation or elevations. Checks are obtained by closing back on the point of beginning or by closing on other points of known elevation. Total error, if within allowable limits, is distributed along the level survey.

Reciprocal leveling, a variant of straight differential leveling, is applied to long distances that cannot be occupied, such as a river crossing or a ravine Here a short backsight to the last turning point and a long foresight (across the inaccessible distance) to a turning point on the opposite side are taken. Then the level is set up on the opposite side so that a long backsight is made to the original point of known elevation and a short foresight is taken to the turning point The average of the two elevation differences thus determined is the actual difference, assuming refraction has remained constant

Trigonometric leveling Elevation differences may also be determined by trigonometric means If the horizontal distance between two points is measured and the vertical angle is measured between the horizontal and a line joining the two points, the elevation difference can be computed

Barometric leveling. Approximate elevation differences may be measured with the aid of a barometer. This method is particularly useful for reconnaissance of substantial areas

Astronomical observations. To determine meridian direction and latitude, observations are made on Polaris, the Sun, or on other stars. Meridian direction is required for direction control purposes, latitude is required as a prerequisite to solar observations, where maps and other available data do not furnish latitude with sufficient precision

Meridian determination. The simplest meridian observation is a sighting on Polaris at elongation, the time when the star is rounding the east or west extremity of its apparent orbit. At these times Polaris appears to move up or down the transit's vertical cross hair and there is ample time for assuring an accurate sighting. An angular correction is applied to determine the direction of sighting which is referenced to a line on the ground The correction value is found in an ephemeris (*see* EPHEMERIS).

The ephemeris also gives the noon declination of the Sun at Greenwich throughout the year Declination is the angle the Sun makes with an equatorial plane. For direct solar observation the direction to the Sun is found by

$$\cos Z = \frac{\sin D}{\cos h \cos L} - \tan h \tan L$$

where Z is the horizontal angle east or west from true north to the Sun; D is declination at the instant of observation; h is the vertical angle to the Sun, corrected for refraction and parallax; and L is the latitude

Latitude determination. Latitude may be observed directly at night by vertical-angle sighting on Polaris at upper or lower culmination (the north or south extremity, respectively, of the star's apparent orbit), and by application of the suitable vertical-angle correction from the ephemeris. In another method, the vertical angle to the Sun's lower edge is observed at noon, the Sun's angular semidiameter is added, and the Sun's declination is subtracted algebraically to yield a net angle, which is subtracted from 90° to give the latitude

Longitude and time observations are made occasionally by noting the passage of the Sun across an established meridian. If Greenwich time is known longitude can be computed

Types of surveys. These include geodetic control surveys, route surveys, construction surveys, and cadastral (property) surveys. *See* TOPOGRAPHIC SURVEYING AND MAPPING.

Geodetic control surveys. Control surveys provide the reference framework for lesser surveys A traverse, with elevations of its points, may be the control for mapping a limited area. The broadest control surveys are the geodetic networks established by the USCGS in the United States and by corresponding agencies in other countries, wherein the horizontal and vertical points are established with first- or second order accuracy.

In geodetic surveys, earth curvature is taken into account Coordinate positions are established in terms of latitude and longitude Geodetic coordinates are convertible into state plane coordinates

The established method of extending horizontal geodetic control is triangulation. Supplementary control is provided by traverse, and trilateration is an alternative technique

In geodetic triangulation, refinements are added to the simple base-line and chain of triangles scheme. The most notable refinement is use of the quadrilateral as a geometric unit (Fig 4) Diagonal directions as well as the sides are observed so that the quadrilateral includes two pairs of overlapping triangles Thus additional angular checks are available and four separate calculations can be made for the entering side length of the next quadrilateral in the ch[illegible] shapes of the

g 4. Triangulation figures

riangles are important, small angles being avoided preserve high strength of figure A strength of figure analysis for the four different computational routes reveals the one that will best carry the length forward.

For precision and economy geodetic triangulation stations are located as far apart as possible, consistent with the requirements of intervisibility. A precise control network for a city may require numerous auxiliary stations within figures

Angles are measured by a precise theodolite (*see* THEODOLITE), all angles around the horizon being observed several separate times for determination of the most probable value of each angle Long sightings usually are made on lighted stations at night.

Because of the earth's shape, the interior angles of large triangles add up to more than 180°. A 75-sq mile triangle for example, has 1 second more than 180° total of interior angles The extra second is called the spherical excess.

Each precisely determined distance between geodetic survey stations is reduced to its sea-level projection, that is, the distance between the stations if they were projected vertically to sea level The sea level projection is a function of the measured distance, its average elevation above sea level, and the mean radius of the earth. The direction between points is given as a geodetic azimuth (degrees clockwise from true south).

Geodetic data bear definite relationships to the established state plane coordinate systems Each system is expressed as a projection of the curved surface on a plane intersecting either two standard parallels (latitudes) or two meridians In both cases the meridian and parallel projections are at right angles to each other At the standard meridians or parallels, scale is exact, elsewhere a known variable scale factor must be applied

Conversions from geodetic data to state plane coordinate data are made with the aid of tables published by the USCGS

Advantages of state plane coordinate surveys include the ability to initiate a long route survey several places at once, and to reestablish obliterated survey points, such as property corners.

The vertical control system of the USCGS consists of a first-order network with supplementary second-order lines In most parts of the United States, bench-mark elevations are available within a few miles of any point.

Leveling procedure is a refinement of that previously described The leveling instrument is built to rigid specifications. Rod graduations are on an invar strip (invar expands and contracts only slightly with temperature changes). Special care is taken to equalize foresights and backsights, to assure stable turning points, and to protect the instrument from minor stresses.

Route surveys Surveys for the design and construction of linear works, such as roads and pipelines, are called route surveys They begin with reconnaissance and continue through preliminary, location, and construction surveys.

Reconnaissance for a new highway, for example, may be accomplished by study of existing maps together with a visual appraisal of field conditions, or it may involve horizontal and vertical field measurements of low order precision Controlling points, such as favorable river and ridge crossings, are found, and a preliminary line is selected This line traditionally is traversed by transit and tape and is profiled that is levels are run along the traverse line to determine its elevations Transverse profiles are made at intervals Structures and natural objects that would affect the final location are fixed by side shots. A side shot may be a direction and distance from a transit point, the intersection of two directions or distances or a perpendicular offset from a line The transverse profiles, called cross sections, may be made by hand level, tape, and level rod

The result of the preliminary survey is a strip topographic map of sufficient precision to permit preliminary design of the final location, including approximate determination of earthwork quantities

The location survey line is conducted with at least third-order precision Traverse procedures are followed, but curves are laid out and stationed, so the result is a staked centerline for the route to be constructed Profile leveling and cross-sectioning for earthwork quantities also are a part of the location survey.

Contract plans normally are based on the location survey. The construction stake-out is usually conducted later but sometimes is carried out at the same time as the location survey. A prerequisite is final grade selection so that cuts and fills can be marked on centerline stakes and slope stakes can be set The slope stake indicates the lateral limits of cut or fill at a given cross section. It is marked with the distance from the centerline and the vertical distance, plus or minus, from existing ground level to the proposed centerline grade. Being at the edge of earthwork, the slope stake supposedly is available for reference throughout construction

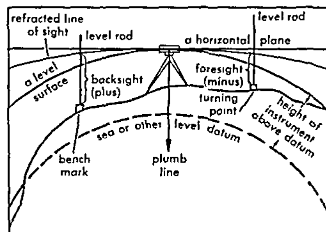


Fig. 3. Differential leveling theory

and effects of earth curvature and refraction of the line of sight. These are illustrated in exaggerated form in Fig. 3.

The foregoing process is duplicated successively as necessary to obtain the desired new elevation or elevations. Checks are obtained by closing back on the point of beginning or by closing on other points of known elevation Total error, if within allowable limits, is distributed along the level survey

Reciprocal leveling, a variant of straight differential leveling, is applied to long distances that cannot be occupied, such as a river crossing or a ravine. Here a short backsight to the last turning point and a long foresight (across the inaccessible distance) to a turning point on the opposite side are taken. Then the level is set up on the opposite side so that a long backsight is made to the original point of known elevation and a short foresight is taken to the turning point The average of the two elevation differences thus determined is the actual difference, assuming refraction has remained constant

Trigonometric leveling. Elevation differences may also be determined by trigonometric means. If the horizontal distance between two points is measured and the vertical angle is measured between the horizontal and a line joining the two points, the elevation difference can be computed.

Barometric leveling Approximate elevation differences may be measured with the aid of a barometer. This method is particularly useful for recon-

latitude is required as a prerequisite to solar observations, where maps and other available data do not furnish latitude with sufficient precision

Meridian determination. The simplest meridian observation is a sighting on Polaris at elongation, the time when the star is rounding the east or west extremity of its apparent orbit. At these times Polaris appears to move up or down the transit's vertical cross hair and there is ample time for assuring an accurate sighting. An angular correction is applied to determine the direction of sighting, which is referenced to a line on the ground The correction value is found in an ephemeris (see EPHEMERIS).

The ephemeris also gives the noon declination of the Sun at Greenwich throughout the year Declination is the angle the Sun makes with an equatorial plane. For direct solar observation the direction to the Sun is found by

$$\cos Z = \frac{\sin D}{\cos h \cos L} - \tan h \tan L$$

where Z is the horizontal angle east or west from true north to the Sun; D is declination at the instant of observation; h is the vertical angle to the Sun, corrected for refraction and parallax; and L is the latitude.

Latitude determination. Latitude may be observed directly at night by vertical-angle sighting on Polaris at upper or lower culmination (the north or south extremity, respectively, of the star's apparent orbit), and by application of the suitable vertical-angle correction from the ephemeris In another method, the vertical angle to the Sun's lower edge is observed at noon, the Sun's angular semidiameter is added, and the Sun's declination is subtracted algebraically to yield a net angle, which is subtracted from 90° to give the latitude

Longitude and time observations are made occasionally by noting the passage of the Sun across an established meridian. If Greenwich time is known, longitude can be computed

Types of surveys. These include geodetic control surveys, route surveys, construction surveys, and cadastral (property) surveys. *See* TOPOGRAPHIC SURVEYING AND MAPPING.

Geodetic control surveys. Control surveys provide the reference framework for lesser surveys A traverse, with elevations of its points, may be the control for mapping a limited area The broadest control surveys are the geodetic networks established by the USCGS in the United States and by corresponding agencies in other countries, wherein the horizontal and vertical points are established with first- or second-order accuracy.

In geodetic surveys, earth curvature is taken into account. Coordinate positions are established in terms of latitude and longitude Geodetic coordinates are convertible into state plane coordinates

The established method of extending horizontal geodetic control is triangulation Supplementary control is provided by traverse, and trilateration is an alternative technique.

In geodetic triangulation, refinements are added to the simple base-line and chain-of-triangles scheme. The most notable refinement is use of the quadrilateral as a geometric unit (Fig 4) Diagonal directions as well as the sides are observed, so that the quadrilateral includes two pairs of overlapping triangles Thus additional angular checks are available and four separate calculations can be made for the entering side length of the next quadrilateral in the ch e shapes of the

Elevation controls are provided by bench marks near the construction area. The foregoing construction techniques are adaptable into industrial plants for building large mock-ups and jigs as well as for the alignment of parts Transits and levels on stable mounts may be used; however, related specialized equipment called optical tooling instruments are more readily applied. These include a transit or level telescope with an optical micrometer on the objective, to move the line of sight to a locus par-

transit base (Fig 6)

Underground surveys Mine and tunnel surveys impose a few modifications on normal surveying techniques: repeated independent measurements are made because normal checks (such as closed traverses) are not available, cross hairs are illuminated because work is performed in relative darkness; vertical tape measurements and trigonometric levels, instead of differential levels, frequently must be relied upon; and in adits and tunnels, survey points are placed overhead, rather than underneath to save them from disturbance by traffic On the mining transit, an auxiliary telescope outside the trunnion bracket facilitates steep sightings.

The most exacting underground survey process is transferring a direction from the surface. In shallow shafts steep (but not vertical) sights may transfer the direction Another technique is to hang two weighted wires down the shaft observe the direction between them on the surface and use this direction as a control below The relative shortness of distance between wires makes the transfer geometrically weak, but procedural care enables satisfactory results

Hydrographic surveys Data for navigation charts and underwater construction are provided by hydro-...ic surveys The horizontal locations of depth ...urements must be referenced to recognizable ...ols. Where the shoreline is visible, it is ...ped and a system of triangulation stations ...tablished on shore Transits at two triangula... stations may be used to observe angles to the ...ding vessel whenever it signals that a depth ...surement is made. A check angle may be ob...ed by sextant from the vessel or a third transit ...shore may provide a check intersection Angular ...ervations also are made from a common center ...nt. In this case two sextants at the sounding ...nt sight two adjacent pairs of shore stations to ...tain the fix.

Hydrographic surveys also determine current ...locities and directions and fluctuations of water ...vel *See* SEA LEVEL FLUCTUATIONS; SUBMARINE ...POGRAPHY.

Cadastral surveys To establish property lines, ...dastral surveys are made. Descriptions based on ...orizontal surveys are essential parts of any docu...ment denoting ownership or conveyance of land ...he basic rule of property lines and corners is

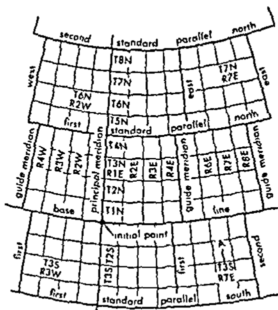


Fig 7. Rectangular land layout (From P. Kissam, Surveying for Civil Engineers, McGraw-Hill, 1956)

that they shall remain in their original positions as established on the ground

The basic rule is important because most land surveys are resurveys. They may follow the original description, but this description is merely an aid to the discovery of the originally established lines and corners Substantial discrepancies frequently are found in original descriptions, because low-order surveying devices such as compasses and chains were used

Surveys in the original 13 states, Tennessee, Kentucky, and a few other areas are conducted on the metes and bounds basis In the so-called public land states and in Texas the basic subdivisions are rectangular.

A metes and bounds survey is a closed traverse around a property Its description identifies a point of beginning, gives the sequence of distances and directions, identifies the angle points, and notes the fact of return to the point of beginning

In the rectangular system, the land parcels of a region are described by their relationship to an initial point In public land states the ...

normally 36 square miles, are designated by their position with respect to the initial point—the number of tiers north or south of a given base line and the number of ranges east or west of the corresponding principal meridian

Within the township each square mile, or section, has a number, from 1 to 36 Sections are subdivided into quarter sections (160 acres) and they may be subdivided further into lesser fractions. North-south section lines are meridians originating at 1-mile intervals along the base line and along standard parallels of latitude (correction lines) generally spaced 24 miles apart The respacing of merid-

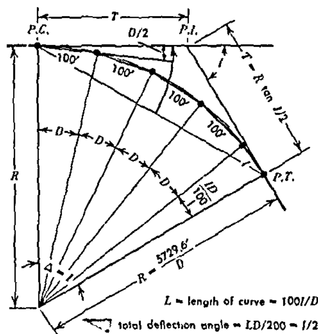


Fig 5. Circular curve theory.

A horizontal curve provides for change in direction for traffic. The directions, called tangents, connect at an angle point, called the point of intersection (*P.I.*). The intersection angle *I* is the deflection angle, right or left, from the forward projection of the tangent into the angle point (Fig. 5).

Any circular curve or arc, connecting two intersecting tangents, has a central angle, equal to the angle *I*, formed by perpendiculars (radii) to the two tangents, one at the point of curvature (*P.C.*) and the other at the point of tangency (*P.T.*). The greater the radius, the more gradual the curve. The sharpness of the curve also is expressed as degree of curvature *D*, the central-angle increment that subtends 100 ft of arc (highway definition) or 100 ft of chord (railway definition). A 1° curve has a radius of approximately 5729.6 ft under either definition.

Circular curves are laid out by the following procedure, assuming *D*, and consequently radius *R*, have been chosen:

1. The tangent distance is measured back from the *P.I.* to establish the *P.C.*

2. The transit is set up at *P.T.* and a series of deflection angles, each equal to *D*/2 is turned in the direction of curvature to establish a point on the curve for each 100 ft of chord or arc. For chords or arcs shorter than 100 ft the turned deflection angle is proportionately smaller.

Vertical curves, connecting different grades usually are parabolic. The parabola has an inherent transition, and it is easy to lay out with tape, level, and level rod after computation of ordinates for vertical-curve grade points along the centerline.

The traditional route-survey procedures described above are economical for narrow routes in moderate terrain. Major dual highways with broad rights of way are located more efficiently on large-scale aerial topographic maps. On such maps, "paper" location surveys can be made, complete with earthwork quantity takeoff pre enough for contract bidding. With this appro it may be feasible to defer all field staking u construction time, the staking points being sele in advance by scaling and computation of n coordinate relationships with points establi during the small amount of ground survey requ to control the photogrammetric survey. See PHO GRAMMETRY.

Construction surveys. Surveys for construc layout establish systems of reference points are not likely to be disturbed by the work. Sl stakes are earthwork references. Buildings, bri abutments, sewers, and many other structures c monly are controlled by batter boards, horizo boards fastened to two uprights. The top of a ter board is set at the elevation of the line to established, and the horizontal position of the is indicated by a mark or nail. Across the site other batter board is set up for the given line an string or wire stretched between is the line line may be a building face at first floor level may be a reference line. In trench work the line tween batter boards may run at a fixed dist above the invert centerline.

In lieu of string or wires, line-of-sight may used. A transit is set up on lines outside the w area and points of the line are sighted as requi A means of locating critical construction poi such as those for anchor bolts, is to compute t positions in a coordinate grid, then to compute rections and distances from a reference point their location by transit sighting and tape meas ment.

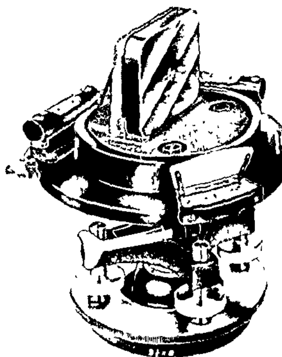


Fig. 6. Autoreflection angle mirror for establish angular lines of sight in optical tooling (Keuffel & Esser Co.)

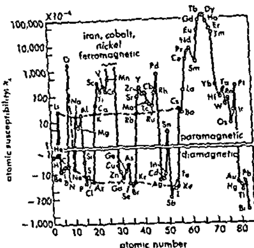


Fig. 1. Atomic susceptibilities of the elements at room temperature Dotted lines connect alkali metals (paramagnetic) and rare gases (diamagnetic) (*After R M Bozorth*)

This behavior is followed in the region well above the ferromagnetic Curie temperature T_c The paramagnetic Curie temperature θ is usually *slightly* greater than the temperature of transition, T_c Comparison of θ and T_c for three ferromagnetic metals is given in the accompanying tabulation.

	Fe	Co	Ni
θ, °K	1093	1428	650
T_c, °K	1043	1393	631

All ferromagnetic materials exhibit paramagnetic behavior above their ferromagnetic Curie points The magnitude of the paramagnetic susceptibility is determined by the Curie constant C which appears in Eq (2) A typical value of C is 0 2°K/cm³ for iron For the theory of ferromagnetic susceptibility, *see* FERROMAGNETISM, *see also* CURIE TEMPERATURE, MAGNETIC, CURIE WEISS LAW

The Curie-Weiss law for the susceptibility above the Curie point is only approximately true. Furthermore the susceptibility of iron, for example, shows discontinuities at its phase changes β–γ and γ–δ at 1183°K and 1673°K, respectively

Below the Curie point the static susceptibility is not usually defined for a ferromagnetic substance, since the ferromagnet may have a finite magnetization in zero applied field

The initial *permeability* is *the slope* of the magnetization curve (B plotted against H) for magnetic field strength $H = 0$ Initial permeabilities vary from almost 300 (platinum-cobalt) to 100,000 (supermalloy)

The frequency dependent permeability is that measured in an alternating magnetic field The experiment ... fields ... initial ... Such experiments give information on the *structure* of ferromagnetic domains and the motion of domain walls

Paramagnetic susceptibility. Most paramagnetic substances at room temperature have a static susceptibility which follows a Langevin-Debye law:

$$\chi = Np^2\mu_B^2/3kT + N\alpha \qquad (3)$$

where N is the number of magnetic dipoles per unit volume, p is the effective magneton number, μ_B is the Bohr magneton, k is Boltzmann's constant, T is the absolute temperature, and α is the temperature-independent contribution of Van Vleck paramagnetism The first term of Eq (3) is referred to as the Curie law because of its $1/T$ dependence. Experimental results are often expressed in terms of what the effective magneton numbers must be in order to account for the variation of χ with $1/T$. The interpretation of the experimental data reveals the nature of the energy levels of the paramagnetic ions, the symmetry of paramagnetic crystals, the effects of crystalline electric fields on the energy levels, and the influence of the paramagnetic ions on one another. *See* PARAMAGNETISM.

The magnitude of molar paramagnetic susceptibilities at room temperature is of the order of 10^{-3} ergs per oersted per mole

Saturation of the paramagnetic susceptibility occurs when a further increase of the applied magnetic field fails to increase the magnetization, because *practically* all the magnetic dipoles are already oriented parallel to the field Saturation will occur either at very strong fields or at very low temperatures, that is, in cases when the *approximation* leading to the Langevin-Debye formula fails Some paramagnetic solids have susceptibilities which follow a Curie-Weiss law rather than a Curie law

Diamagnetic susceptibility. The susceptibility of diamagnetic materials is negative, since a diamagnetic *substance* is magnetized in a direction opposite to that of the applied *magnetic* field. The diamagnetic susceptibility is independent of temperature

Diamagnetic susceptibility depends upon the distribution of electronic charge in an atom and upon the energy levels Interpretation of the experimental data reveals the nature of the atomic wave functions of the atom in question. For the *theory of* diamagnetic *susceptibility, and a listing of values* ... above the ... from anti- ... to paramagnetic behavior, follows a Curie-Weiss law with a negative paramagnetic Curie temperature $-\theta$:

$$\chi = C/(T + \theta)$$

The *Néel* temperature is always somewhat less than θ

The antiferromagnetic susceptibility is a maximum at the Néel temperature. Below the Néel temperature, it behaves in *a way determined* by the an-

ional lines every 24 miles limits the effects of meridian convergence on the size of sections.

Any parcel, whether it is described in rectangular terms or by metes and bounds, should be marked at its corners. Corner markers and other physical evidence of parcel boundaries take precedence over written descriptions and statements of areas. Resurveys usually are simple if the corners can be found. They become difficult where the corners are obliterated or lost.

An obliterated corner is one for which visible evidence of the previous surveyor's work has disappeared but whose original position can be established from other physical evidence and testimony, beyond reasonable doubt.

A corner is deemed lost when no sufficient evidence of its position can be found. Restoration requires a faithful rerun of the recorded original survey lines from adjacent points, distance discrepancies being adjusted proportionately. [R.H.DO]

Bibliography: R. E. Davis and F S. Foote, *Surveying: Theory and Practice*, 4th ed, 1953; P Kissam, *Surveying for Civil Engineers*, 1956; H. Rubey, G. E. Lommel and M W. Todd, *Engineering Surveys*, 1958.

Susceptance

The imaginary part of the complex representation for the admittance Y of a circuit.

$$Y = G \pm jB$$

where G is the real part, called the conductance, and B is the susceptance. Since $Y = 1/Z = 1/(R + jX)$ where X is the total reactance, $X_L - X_C$, and R is the resistance, then

$$Y = \frac{R}{R^2 + X^2} - j\frac{X}{R^2 + X^2}$$

and susceptance $B = X/(R^2 + X^2)$. This is the general expression for susceptance which shows that susceptance is a function involving both resistance and reactance

If resistance is negligible, then $B = X/X^2 = 1/X$ or the reciprocal of the reactance This is called simple susceptance and is correct only where the impedance contains no resistance. For simple susceptances

$$\text{Inductive susceptance } B_L = 1/jX_L = -jB_L \quad \text{mhos}$$

$$\text{Capacitive susceptance } B_C = 1/-jX_C = jB_C \quad \text{mhos}$$

These functions find application chiefly in computation of parallel circuits. *See* ADMITTANCE, ALTERNATING-CURRENT CIRCUIT THEORY. [B.L.R.]

Susceptibility, electric

A dimensionless parameter measuring the ease of polarization of a dielectric. The susceptibility χ is equal to the ratio of polarization P to the product of electric field strength E and vacuum permittivity ϵ_0:

$$\chi = \mathrm{P}/\epsilon_0\mathrm{E}$$

($\epsilon_0 = 1$ in cgs electrostatic units, and 8.854×10^{-1} farad/m in mks units.)

The electric susceptibility is related to the dielectric constant κ' by the equation:

$$\chi = (\kappa' - 1)/\gamma$$

where γ is a geometrical factor equal to 4π or 1 in the cgs or mks systems, respectively. This ambiguity in the definition of χ is unfortunate and must be considered in evaluating published data. *See* ELECTRICAL UNITS.

The electric susceptibility can be related to the polarizability α by expressing the polarization in terms of molecular parameters Thus:

$$\mathbf{P} = N\langle\mathbf{\mu}\rangle_{avg} = N\alpha\mathrm{E}_L \quad \text{and} \quad \chi = \frac{N\alpha\mathrm{E}_L}{\epsilon_0\mathrm{E}}$$

where N is the number of molecules per unit volume, $\langle\mathbf{\mu}\rangle_{avg}$ is their average dipole moment, and E_L is the local electric field strength at a molecular site. At low concentrations, E_L approaches E, and the susceptibility is proportional to the concentration N.

For a discussion of the properties and measurement of electric susceptibilty, *see* DIELECTRIC CONSTANT; POLARIZATION (DIELECTRICS). [R.D.W]

Susceptibility, magnetic

The magnetization of a material per unit applied field. It describes the magnetic response of a substance to an applied magnetic field. If M is the magnetization and H the applied magnetic field, then the magnetic susceptibility, denoted by χ, is

$$\chi = M/H \tag{1}$$

In the case that M is not parallel to H, χ is a tensor Otherwise, it is a simple number. For an elementary discussion of M and H, *see* MAGNETISM; *see also* MAGNETIZATION. For a crystalline material, χ may depend upon the direction of H with respect to the axes of the crystal because of anisotropic effects

The magnetic susceptibility is expressed in a variety of ways: per gram, per atom, per unit volume, and per mole In this article, electromagnetic units are used. In Eq. (1), the units of χ are ergs per oersted per unit volume. Figure 1 shows the atomic susceptibilities χ_A (ergs per oersted per atom) of the elements. The static susceptibility is measured in constant applied magnetic fields. The frequency-dependent susceptibility is measured in alternating magnetic fields It is usually a complex quantity in which both the real and imaginary

The general be-

rromagnetic materials above the Curie temperature follows the Curie-Weiss law

$$\chi = \frac{C}{T - \theta} \tag{2}$$

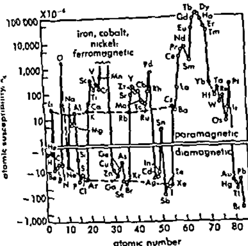


Fig. 1 Atomic susceptibilities of the elements at room temperature. Dotted lines connect alkali metals (paramagnetic) and rare gases (diamagnetic). (After R. M. Bozorth)

This behavior is followed in the region well above the ferromagnetic Curie temperature T_c. The paramagnetic Curie temperature θ is usually slightly greater than the temperature of transition, T_c. Comparison of θ and T_c for three ferromagnetic metals is given in the accompanying tabulation.

	Fe	Co	Ni
θ, °K	1093	1428	650
T_c, °K	1043	1393	631

All ferromagnetic materials exhibit paramagnetic behavior above their ferromagnetic Curie points. The magnitude of the paramagnetic susceptibility is determined by the *Curie constant C* which appears in Eq. (2) . . .

0.2°K/ct

susceptil

TEMPERA

The Curie-Weiss law for the *susceptibility* above the Curie point is only approximately true. Furthermore, the susceptibility of iron, *for example*, shows discontinuities at its phase changes β–γ and γ–δ at 1183°K and 1673°K, respectively.

Below the Curie point, the static susceptibility is not usually *defined* for a ferromagnetic substance, since the ferromagnet may have a finite magnetization in zero applied field.

The initial permeability is the slope of the magnetization curve (B plotted against H) for magnetic field strength $H = 0$. Initial permeabilities vary from almost 300 (platinum-cobalt) to 100,000 (supermalloy).

The frequency-dependent permeability is *that* measured in an alternating magnetic field. The experiments are usually carried out in small magnetic fields so that it is the frequency dependence of the initial permeability which is measured. Such experiments give information on the structure of ferromagnetic domains and the motion of domain walls.

Paramagnetic susceptibility. Most paramagnetic substances at room temperature have a static susceptibility which follows a Langevin Debye law:

$$\chi = Np^2\mu_B^2/3kT + N\alpha \qquad (3)$$

where N is the number of magnetic dipoles per unit volume, p is the effective magneton number, μ_B is the Bohr magneton, k is Boltzmann's constant, T is the absolute temperature, and α is the temperature-independent contribution of Van Vleck paramagnetism. The first term of Eq. (3) is referred to as the Curie law because of its $1/T$ dependence. Experimental results are often expressed in terms of what the effective magneton numbers must be in order to account for the variation of χ with $1/T$. The interpretation of the experimental data reveals the nature of the energy levels of the paramagnetic ions, the symmetry of paramagnetic crystals, the effects of crystalline electric fields on the energy levels, and the influence of the paramagnetic ions on one another. *See* PARAMAGNETISM.

The magnitude of molar paramagnetic susceptibilities at room temperature is of the order of 10^{-4} *ergs per oersted per mole*.

Saturation of the paramagnetic susceptibility occurs when a further increase of the applied magnetic field fails to increase the magnetization, because practically all the magnetic dipoles are already oriented parallel to the field. Saturation will occur either at very strong fields or at very low temperatures; that is, in cases when the approximation leading to the Langevin-Debye formula fails. Some paramagnetic solids have susceptibilities which follow a Curie-Weiss law rather than a Curie law.

Diamagnetic susceptibility. The susceptibility of diamagnetic materials is negative, since a *diamagnetic substance* is magnetized in a direction opposite to that of the applied magnetic field. The diamagnetic susceptibility is independent of temperature.

Diamagnetic susceptibility depends upon the distribution of electronic charge in an atom and *upon* the energy levels. Interpretation of the experimental data reveals the nature of the atomic wave *functions* of the atom in question. For the theory of diamagnet . . .

. . . antiferromagnetic materials above the Néel point, which marks the transition from antiferromagnetic to paramagnetic behavior, follows a Curie-Weiss law with a negative paramagnetic Curie temperature $-\theta$:

$$\chi = C/(T + \theta)$$

The Néel temperature is always somewhat less than θ.

The antiferromagnetic susceptibility is a maximum at the *Néel temperature*. Below the Néel temperature, it behaves in a way determined by the an-

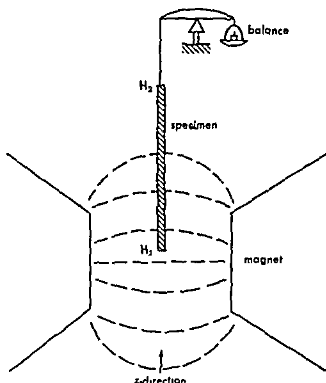


Fig 2. Gouy balance method for measurement of magnetic susceptibilities. (After C. Kittel, Introduction to Solid State Physics, 2d ed., Wiley, 1956)

gle between the field direction and the crystal axes —it usually decreases as the temperature lowers. For the theory of antiferromagnetic susceptibility, *see* ANTIFERROMAGNETISM

Measurement of susceptibility. Magnetic susceptibilities may be measured by several methods, depending upon the quantity sought. Ferromagnetic static permeabilities may be measured by the Rowland ballistic method In this technique, the specimen is cut in the shape of a ring, and a magnetic field is applied by means of a primary winding on the ring. A secondary winding is connected to a ballistic galvanometer. The current through the primary is changed suddenly. The resulting abrupt change in magnetic field H causes a change in magnetic induction B $(B = H + 4\pi M)$ in the specimen which in turn induces a voltage in the secondary. Thus, the galvanometer suffers a deflection proportional to the change in B One starts with a known value of B (usually zero) and plots the magnetization curve in this way.

Paramagnetic and diamagnetic susceptibilities may be measured by the balance method A magnetized substance in an inhomogeneous magnetic field experiences a force

$$F = \tfrac{1}{2} \text{ grad } \int M \cdot H \, dV$$

The integral is over the volume of the specimen The susceptibility is $\chi = M/H$. Therefore

$$F = \tfrac{1}{2}\chi \text{ grad } \int H^2 \, dV$$

If the magnetic field now varies only in one direction, say the z-direction, then

$$F = \tfrac{1}{2}\chi A \int \frac{d}{dz} (H^2) \, dz = \tfrac{1}{2}\chi A(H_1^2 - H_2^2)$$

where the specimen is in the shape of a rod of cross-sectional area A and is suspended in the z direction in the inhomogeneous field (see Fig 2) In the figure, H_1 and H_2 are the values of the magnetic field strength at the two ends of the rod The force is measured, and when H_1 and H_2 (or H and dH/dz) are known, the susceptibility can be determined. This method is known as the Gouy balance method. Note that the total magnetic susceptibility is measured, both para- and diamagnetic contributions.

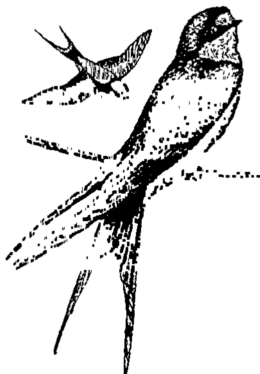
The paramagnetic susceptibility may sometimes be measured separately by spin-resonance techniques. This is especially true of the Pauli paramagnetic susceptibility of conduction electrons in alkali metals. *See* MAGNETIC RESONANCE.

[E.A., F.K.]

Bibliography: C. Kittel, *Introduction to Solid State Physics*, 2d ed., 1956; J. H. Van Vleck, *The Theory of Electric and Magnetic Susceptibilities*, 1932.

Swallow

Any of about 75 species of the cosmopolitan family Hirundinidae, characterized by modifications for strong flight and feeding while in flight, including long pointed wings, weak short legs, and a short triangular beak. There are six genera and nine species in the United States Other than the purple martin, the best known species is the barn swallow, *Hirundo rustica* This graceful, long-tailed swallow breeds throughout the United States, except in pen



The barn swallow, Hirundo rustica; length to 7¾ in. (From E L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

insular Florida, although it is not nearly as common in the South as elsewhere. Steel blue and cinnamon in color, it has a long forked tail. It nests habitually near human habitation.

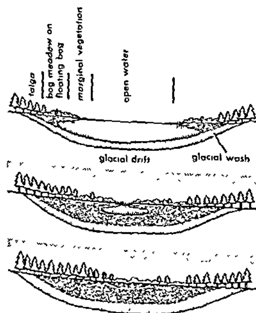
Most beautiful of the swallows is the tree swallow *Tachycineta bicolor*, marked with immaculate white underparts, with a back which varies from green black to steely blue-black. This is a bird of the ponds, wet meadows, and marshes. It collects in large flocks in late summer. The bank swallow, *Riparia riparia*, is a wide-ranging species occurring throughout the Holarctic region and Africa as a nesting bird. *See* MARTIN; PASSERIFORMES.

[J.D.B.]

Swamp, marsh, and bog

Wet flatlands where mesophytic vegetation is areally more important than open water, are commonly developed in filled lakes, glacial pits and potholes, or poorly drained coastal or flood plains. Swamp is a term usually applied to a wet land where trees and shrubs are an important part of the vegetative association, and bog implies lack of solid foundation. Some bogs consist of a thick zone of vegetation floating on water.

Unique plant associations characterize wet lands in various climates and exhibit marked zonation characteristics around the edge in response to different thicknesses of the saturated zone above the firm base of soil material. Coastal marshes covered with vegetation adapted to saline water are common on all continents. Presumably many of these had their origin in recent inundation due to post-Pleistocene rise in sea level.



Cross-section diagram illustrating progressive filling by vegetation of a pit lake in recently glaciated terrain. (After C. A. Davis)

The total area covered by these physiographic features is not accurately known, but particularly in glaciated regions many hundreds of square miles are covered by marsh.

[L.B.L.]

Swan

Any of eight members of the cosmopolitan subfamily Cygninae, of the waterfowl family Anatidae. There are two species native to the United States and a third has been introduced. Swans are large



The domestic swan, *Cygnus olor*; length 60 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

waterfowl, characterized by very long necks. In flight they may form V-shaped [illegible]

swimmers and will frequently swim away when molested rather than take flight. A long run is necessary for them to become airborne. All the United States swans are white when mature, and may be distinguished from geese and other large water birds by the absence of black on their wing tips. The brown-tinged young are called cygnets. Swans feed primarily upon vegetable matter, but also eat mollusks and crustaceans. *See* ANSERIFORMES.

[J.D.B.]

Swanscombe man

A partial skull vault recovered from the 100-ft terrace gravels of the Thames River, in the Barnfield Pit, Swanscombe, Kent, England. The occipital [illegible] in [illegible] 19 [illegible] though separated by distances of 8 yards, the bones fit one another perfectly and all derive from the same layer of the middle gravels, which contains Middle Acheulian hand-axes and is dated by fossils to the latter part of the Second, or Great, Interglacial.

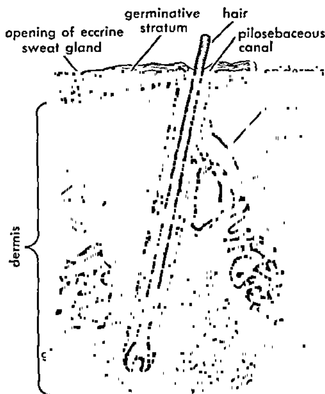
The skull is morphologically close to *Homo sapiens* in having a vertical temporal region and a rounded occipital profile which distinguish it from most types of fossil man. Special features are greater thickness and greater occipital breadth than modern man. Brain volume is estimated at approximately 1300 cm^3. Since the skull lacks a frontal region, face, teeth, and mandible, no final assessment of the fossil's position is possible. It may be closely related to Steinheim man. The bones

are at the British Museum (Natural History). *See* FOSSIL MAN. [W.W.H.]

Bibliography: Swanscombe Committee, Roy. Anthrop. Inst., Report on the Swanscombe skull, *J. Roy. Anthrop. Inst.*, 68:17–98, 1938

Sweat gland

A coiled, tubular gland found in mammals There are two kinds, merocrine or eccrine, and apocrine. The latter are generally associated with hair follicles (see illustration). Merocrine glands are distributed extensively over the body in the human, whereas the apocrine variety is restricted to the scalp, nipples, axilla, external auditory meatus, external genitals, and perianal areas Apocrine sweat glands are more numerous in mammals, with the exception of the chimpanzee and human in which the merocrine variety predominates The mammary glands probably represent modified apocrine sweat glands which grow inward and increase in complexity. In association with adipose tissue, they eventually form pendant structures, the mammae, which project outward from the general contour of the skin's surface. The secretion process is apocrine with a considerable portion of the cell being discharged. The discharged portion of such a gland cell disintegrates to free fat droplets and albuminous substances (*see* LACTATION). A mammary gland is complex and represents an association of lobes. Each lobe contains a compound alveolar (acinous) gland with a separate lactiferous duct which opens on the nipple in the human. The glands of Moll associated with the eyelashes are relatively large modified apocrine glands as are the ceruminous or wax glands in the external auditory meatus. The anal sacs of the skunk presumably are apocrine glands modified by the addition of muscle fibers from the levator ani muscle which enables the pungent contents to be ejected with force *See* EPITHELIUM; GLAND; MAMMARY GLAND [D.E.N.]



Human skin. Eccrine and apocrine sweat glands and sebaceous gland.

Sweep generator

An electronic circuit that generates a voltage or current, usually recurrent, as a prescribed function of time. The resulting waveform is used as a time base to be applied to the deflection system of an electron-beam device, such as a cathode-ray tube Sweep generators are classified as linear, circular, rotating radial, or hyperbolic.

Linear sweep generation. A linear sweep generator provides a current or voltage that is a linear function of time. The waveform is usually recurrent at uniform periods of time (*see* SAW-TOOTH WAVE) as shown in Fig. 1. The deflection of the electron beam in a cathode-ray tube of the type normally used in the cathode-ray oscilloscope may be expressed as

$$d = kv_d$$

where k is a constant which is inversely proportional to the beam-accelerating potential, and v_d is the potential applied between the deflection plates. Thus, if the sweep waveform applied to the plates has the form $v_d = k_2t$ at repeating intervals, the deflection of the beam between the plates will be a linear function of time, recurrent at the same rate.

If the deflection system in a cathode ray device is electromagnetic, the deflection of the beam is proportional to the transverse magnetic flux density in the deflection region. In this case, the deflection is approximately proportional to the current in the deflection coils, external to the tube, which produce the magnetic field The sweep waveform is a waveform of current rather than voltage *See* OSCILLOSCOPE, CATHODE-RAY

The well-known raster scan of the television system is produced if simultaneously linear recurrent sweep waveforms are applied to the horizontal and vertical deflection systems of the cathode-ray device, and the vertical period is many times longer than the horizontal period, as shown in Fig. 2 Thus a number of equally spaced, nearly horizontal scans are produced, the number being the ratio of the horizontal sweep period to the vertical sweep period. In the cathode-ray oscilloscope, a desired time-varying function may be visually presented by applying a linear voltage sweep waveform to the



Fig. 1. Saw-tooth wave.

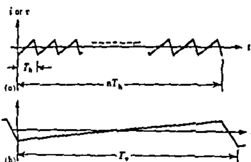


Fig 2 Waveforms producing the raster scan. (a) Horizontal sweep (b) Vertical sweep.

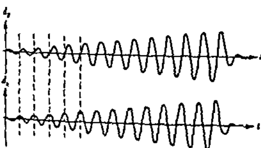


Fig 3 Waveforms producing spiral scan

horizontal deflection system and the function to be plotted to the vertical system Where the raster scan is used, as in the television receiver, the information to be plotted usually appears on the viewing screen as intensity variations produced by modulation of the current in the cathode-ray tube beam *See* TELEVISION SCANNING.

Circular sweep generation. A circular sweep is generated by applying a constant-frequency sinusoid to the horizontal component of a deflection system and another of the same frequency but shifted in phase by 90° to the vertical component If the horizontal component of deflection d_x is $A \cos \omega t$, and the vertical component d_y is $A \sin \omega t$, the total deflection d is the vector sum $\sqrt{d_x^2 + d_y^2}$, which equals A The deflection path is circular at a constant radius A The angular position θ, with respect to the horizontal, is given by $\tan \theta = \tan \omega t$ The angular rate of change of the deflected beam is constant

A spiral scan is generated if, in addition to the sinusoidal x and y components, a linear saw-tooth waveform ($r = kt$) is used to modulate equally both the horizontal and voltage deflection waveform as shown in Fig. 3

Thus

$$d = \sqrt{A^2k^2t^2 \cos^2 \omega t + A^2k^2t^2 \sin^2 \omega t}$$

or

$$d = Akt$$

If the active period of the saw tooth is n times the period of the sinusoid, the spiral will have n revolutions for each period of the saw tooth Circular and spiral sweeps are useful in special forms of oscilloscopes where a long time base is desirable

Rotating radial sweep generation. A rotating radial sweep may be generated by applying a linear sawtooth of current to the deflection system of a magnetically deflected cathode-ray device and rotating the deflection coil producing the magnetic field at a constant angular velocity. Such a sweep may also be generated by using a fixed position deflection system having separate x and y components of deflection. The same combination of linear and sinusoidal modulation as that of the spiral sweep may be used, except that the period of the linear modulation component to the x-y deflection system must be short compared to the period of the angular deflection As indicated in Fig. 4, such a combination of sweeps produces one complete radial line for a negligible change in angle. The rotating radial sweep is most widely used in radar display systems, where the angle of rotation is made synchronous with the scan angle of a radar antenna The linear radial sweep is proportional to radar range. Thus intensity modulation on a display cathode-ray tube having such sweeps will present a true position of individual radar returns. *See* RADAR.

Hyperbolic sweep generation. There are applications, particularly in airborne radar systems, where hyperbolic saw-tooth sweeps are preferred to linear saw-tooth sweeps Such a hyperbolic sweep, shown in Fig 5, may be given by the equation

$$d = k\sqrt{t^2 - T_0^2}$$

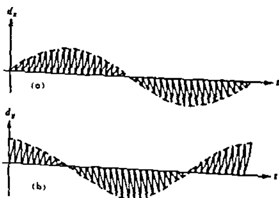


Fig 4 Components of rotating radial sweep (a) Horizontal sweep, (b) Vertical sweep

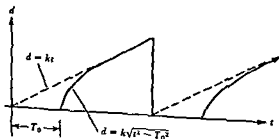


Fig 5 Hyperbolic sweep.

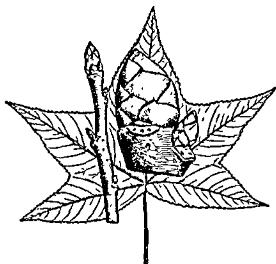
Such a sweep is asymptotic to a linear sweep for large values of t, but is delayed with respect to the beginning of it by a time, T_0. Such a hyperbolic sweep used in a rotating radial sweep system provides for true ground range radar mapping in an airborne system. Here the time T_0 is made equal to h/c, where h is the height of the aircraft, c is the velocity of propagation of electromagnetic energy. The factor k is a constant which represents the scale factor of the display.

Hyperbolic sweeps may be generated as a modification of the type of circuitry used in the generation of saw-tooth sweep waveforms. [G M.G.]

Bibliography: G. M. Glasford, *Fundamentals of Television Engineering*, 1955; T. Soller, M. A. Starr, and G. E. Valley, Jr., *Cathode-ray Tube Displays*, 1948

Sweetgum

This tree, *Liquidambar styraciflua*, also called bilsted or redgum, is a deciduous tree of the southeastern United States. It is found northward as far as southwestern Connecticut, and also grows in Central America. The tree is commonly 120 ft in height and 1½–3 ft in diameter, but individual trees may exceed these dimensions. Sweetgum is readily distinguished by its five-lobed, or star-shaped, leaves and by the corky wings or ridges usually developed on the twigs The erect trunk is a dark gray, but the branches are lighter in color. In winter the persistent, spiny seedballs are an excellent diagnostic feature. The hard, close-grained wood is light brown tinged with red, and has a satiny luster and attractive grain Because of its tendency to warp and twist, it was long considered to be of inferior quality, but technical processing has largely overcome these difficulties. Also, because of the prejudice against gum wood, it is sometimes marketed as satin or Circassian walnut, or as hazelwood.

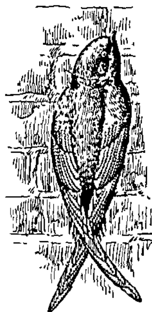


Sweetgum, *Liquidambar styraciflua*. (From A. H. Graves, *Illustrated Guide to Trees and Shrubs*, rev. ed, Harper, 1956)

The saw-timber stand of sweetgum in the United States is estimated at 26,000,000,000 board feet (bd ft). The annual commercial production of 600,000,000–900,000,000 bd ft comes mainly from the South Atlantic and Gulf Coastal Plains and from the lower Mississippi Valley. It is used for furniture, interior trim, railroad ties, cigar boxes, crates, flooring, barrels, woodenware, and wood pulp, and it is one of the most important materials for plywood manufacture. Sweetgum is one of the most desirable ornamental trees, chiefly because of its brilliant autumn coloration See FOREST AND FORESTRY; TREE. [A H C]

Swift

Any of about 75 species of the family Apodidae, an almost cosmopolitan family, characterized by long narrow wings and short tails. Swifts have short, flat bills, large gapes, and small, weak feet All these traits are features of birds that feed while in rapid flight There are four species in the United States East of the Rocky Mountains the chimney swift, *Chaetura pelagica*, is one of the more common summer residents, and always one of the last of the purely insectivorous birds to go south in the fall Swifts have a fluttering, batlike flight quite unlike the easy gliding flight of the swallows. *See* APODIFORMES [J D B]



The chimney swift, *Chaetura pelagica*; length to 5½ in. (From E L Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Swim bladder

The swim bladder is a gas-filled cavity within the body of fishes which serves to equilibrate the density or weight of the body to that of the surrounding water. Most teleosts (bony fishes) possess this organ, although it may be vestigial or absent in bottom-living forms and in sea forms found at abyssal depths Anatomically, there are two types of bladder, the completely enclosed, or physocli-

tous, and those retaining a duct to the digestive tract, or physostomatous. The shape of the bladder is usually ovoid, but in some fish it may be elongate or have a variety of compartmentalizations and protrusions (Fig. 1). The perch, for example, has an anterior and posterior compartment connected by an oval window The whiting has an elongate bladder, extending back into the tail region, with earlike protrusions folded dorsally into spaces between the ribs.

Morphology. Many fish do not retain the swim bladder to gullet passageway. The result is a closed type swim bladder. This is more complicated structurally, and the ability to secrete gases to an amazing pressure, in a manner not yet fully investigated, is characteristic It is this type that has received the most intensive investigation

The envelope or shell of the swim bladder is much the same in both types, being a closely knit, fibrous, tough, smoothly lined membrane However, the closed type is characterized by two additional well vascularized structures, the red gland located on the external surface of the bladder and the secretory epithelium located within Typically, the blood approaches the red gland in one large artery which breaks down into many small vessels

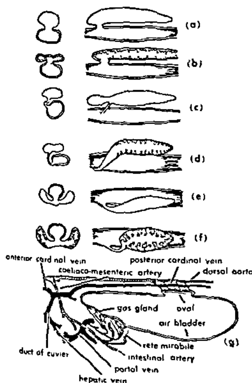


Fig 1 Various types of swim bladders in fish, showing the structures in cross section and longitudinal section. (a) Sturgeon and many teleosts (b) *Lepidosteus* and *Amia* (c) *Erythrinus* (d) *Ceratodus* (e) *Polypterus* (f) *Protopterus* and tetrapods. (g) Physoclistous swim bladder of a teleost (From O E Nelsen, *Comparative Embryology of the Vertebrates*, Blakiston–McGraw-Hill, 1953)

These run parallel to each other in a precise geometric arrangement within the gland. The vessels, on leaving the gland, fuse into several large sinusoidal-like vessels. These penetrate the membrane and immediately branch, in an orderly fashion, to supply a patch of glandular appearing cells which are located on the inside of the bladder membrane.

Blood from the secretory epithelium is returned to collecting vessels, which mesh with the incoming ones. The blood is decanted immediately into another set of parallel vessels within the red gland which are exactly interdigitated with the ones mentioned above. The double set of parallel vessels, half conducting blood one way and half the other way, is known anatomically as a double rete mirabile. It serves as an interchange mechanism, allowing the two blood streams to exchange diffusible substances. The physical principle which chemists now use to great advantage in the countercurrent separation apparatus is the basis. The blood finally leaves the tightly bundled rete by way of one major vessel (Fig 2).

Physiology. Embryologically, the swim bladder is formed as an outpouching from the developing digestive tract, much as the lungs are formed in higher animals In some fish, a permanent connection to the gullet is retained and volume is regulated by periodic swallowing and regurgitation of air. This observation has led to the suggestion that this structure has a respiratory function in addition to the primary one of hydrostatic balance. The theory has not been substantiated because, contrary to the picture in all lungs and gills, there is no well developed capillary bed for respiratory exchange in the open-type bladder. More conclusively, analyses of gases, under experimental conditions, reveal only gradual changes such as would be expected in any body cavity. The lung fishes may be an exception

There is evidence for one other function, phonation For example, the toad fish, though a bottom-living form, possesses a large and tightly filled swim bladder Adjacent to it is a set of muscles that can vibrate the bladder wall as a sounding board and produce a grating croak which is audible long distances through water

Accurately speaking, the swim bladder, as a hydrostatic organ, is not limited to simple equilibration of the density of the fish to the density of the surrounding water. The organ is under the control of the fish and can be used as an active aid in maintaining a desired position. For example, fish induced to swim upward in a constantly descending current will inflate the bladder over its normal limits in order to obtain more buoyancy and *vice versa*. *See* EQUILIBRIUM, BIOLOGICAL; SENSE ORGAN.

If a closed-type swim bladder is deflated by means of a hypodermic needle, it reinflates with a mixture primarily of oxygen (up to 95% having been observed), some carbon dioxide, and some nitrogen. After the original volume is restored, the ratio of gases gradually changes to more nearly

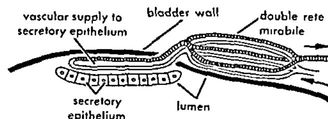


Fig 2. Swim bladder. Schematic diagram of the gas-secreting mechanism.

that of other parts of the body and surroundings where nitrogen is predominant.

Gas production. There are two distinct schools of thought regarding the production of these gases. One is based on the well-known Bohr effect, whereby blood dissociates oxygen and carbon dioxide under acidic conditions and does the reverse under alkaline conditions. The observed ratio of released gases supports this theory in a general way in that oxygen and carbon dioxide are the predominant gases released. Further, it has been observed by histochemical methods that glycogen disappears from the secretory epithelium during active inflation. Manometric respiratory tests show the epithelium to be peculiar in that the metabolism of the glycogen is not completed to the final end product, water and carbon dioxide Instead, it stops at about the lactic acid level and therefore might serve as the acidification mechanism to obtain the Bohr effect Under these circumstances the rete mirabile exchange mechanism might protect the general circulation against an excessive influx of lactic acid and enhance the localized acidic condition of the secretory epithelium.

There is one major flaw to the above analysis It cannot explain the release of gases at pressures up to a hundred times that attributable to an unassisted Bohr effect, which in final analysis is dependent on physical diffusion gradients. Pressure from water increases rapidly with depth and it quickly exceeds the diffusion pressure of dissolved gases which remains relatively constant with depth. Fish that live habitually below 25 ft and do not come periodically to shallow depths where the bladder could be refilled, must inflate the bladder by an active secretion of gases that uses either (1) a metabolic energy mechanism or (2) a physical trick to supplement the Bohr effect

Cellular transport. Those favoring a specific cellular transport mechanism, which utilizes metabolic energy to bind a gas chemically at one point and release it against a pressure gradient at another, base their claims primarily on the necessity of such a mechanism to account for observed pressures. Since all three gases, oxygen, carbon dioxide, and nitrogen, have been observed at pressures above those accountable by the surroundings, this would suggest transport of all three. In such discussions, references are usually made to an interesting parallel situation wherein the Portuguese man-of-war is capable of concentrating the rare gas argon in its float against ambient pressure gradients. However, as yet, no enzymatic or other energy consuming cellular transportation mechanism has been demonstrated for any of the gases concerned The possibility of such a mechanism should not be excluded, but it should also be kept in mind that, because of chemical specificities, a separate enzymatic system would be required for each of the observed gases, oxygen, carbon dioxide, and nitrogen, and probably some of the rarer gases. Transport of the relatively inert nitrogen would be most surprising.

Method of gas analysis. Tests with radioisotopically labeled oxygen has shown that the oxygen released in the swim bladder is identical to the oxygen gas that is in solution in the water about the fish. That is, the swim-bladder oxygen is not derived by a metabolic breakdown of some other substance in the body nor is there an exchange of dissolved oxygen with oxygen already bound in body compounds. Further, isotopic studies show that the molecular oxygen dissolved in the surrounding water eventually reaches the air bladder as the same species of molecule; that is, molecular rather than atomic oxygen is transported. This tends to negate the idea of a chemical or metabolic mechanism of transport.

Theory of gaseous transport. One of the more promising and intriguing theories postulates that oxygen is the only gas transported by a metabolic energy mechanism, while the other gases are introduced by a physical principle based on the diffusion of gases into very small bubbles. If a cellular structure capable of releasing small bubbles of oxygen and accompanying gases can be demonstrated, then this theory may well prove to be the solution of the dilemma.

The theory mentioned in the preceding paragraph receives support by observations, dating back more than 100 years, that the primary gas to refill the closed-type air bladder, after deflation by puncture, is oxygen It has further been hypothesized that once the air-bladder volume has been established, oxygen is selectively reabsorbed, leaving the more "inert" gases The latter thought would presuppose a mechanism to protect or maintain the status quo of gases in the air bladder, except the one actively transported Such does indeed exist in the rete mirabile which was mentioned above as an interchange mechanism. This is based on the countercurrent principle which could protect the general circulation against any acidic dissociative component released by the secretory epithelium By the same token the structure would tend to minimize the loss of any other diffusible substance from the swim bladder Very significant in this connection is the fact that fish living at the greater depths possess the longer retes

Inflation mechanisms It can be said that in the absence of an unusual physical phenomenon, an energy mechanism is involved in the inflation of the closed-type air bladders of fish found at any depth other than very shallow waters. Whether this mechanism is some patient Maxwell's demon or a more

prosaic enzymatic system has not been proven. Whatever the mechanism, evidence indicates that it primarily involves oxygen The other gases very probably are moved passively by means of diffusion gradients

One further variant might be appended to the above discussion. Some fish, as noted at the beginning, have bladders separated into two compartments by an oval window There is good evidence that the opening and the closure of the oval window control the inflation and deflation of the bladder It is hypothesized that constant secretion of gas occurs in one compartment (usually the larger one) and constant absorption in the other. The final status of the bladder is determined by the sphincterlike activity of the oval window, the musculature of which is under the nervous control of the fish. *See* PHONORECEPTION. [D E.CO.]

Bibliography: D E Copeland, The histophysiology of the teleostean physoclistous swimbladder, *J. Cellular and Comp Physiol.*, 40(2):317–335, 1952, W. Jacobs, Untersuchungen zur Physiologie der Schwimmblase der Fische Über die Gassekretion in der Schwimmblase von Physoclisten, *Z. vergleich Physiol*, 11(4):565–629, 1930; J. Ledebur, Über die Sekretion und Resorption von Gasen in der Fischschwimmblase, *Biol Rev*, 12:217–244, 1937, M. Saupe, Anatomie und Histologie der Schwimmblase des Flussbarsches (Perca fluviatilis) mit besonderer Berücksichtigung des Ovals, *Z Zellforsch U Mikroskop Anat.*, 30(1):1–35, 1940, P F. Scholander, L. van Dam and T Enns, Nitrogen secretion in the swimbladder of whitefish *Science*, 123 59–60, 1956; J B Wittenberg, The secretion of inert gas into the swimbladder of fish. *J. Gen Physiol.*, 41(4):783–804, 1958, W N F Woodland, On the structure and function of the gas gland and rete mirabilia associated with the gas bladder of some teleostean fishes with notes on the teleost pancreas, *Proc Zool Soc* London, pp 183–248, 1911

Swine production

The production of swine is an agricultural business which is usually located in proximity to sources of high energy feedstuffs Particularly in the United States, the geographical distribution of swine production is closely related to corn and sorghum production. Swine, as nonruminants, utilize large quantities of concentrate feeds, for instance, it is estimated that 45% of the corn grain is fed to swine. However, swine utilize only limited quantities of roughage in the diet

Swine products. The primary products of swine are pork and lard, hides, and innumerable byproducts of a pharmaceutical nature Pork and lard supply approximately 15% of the total calories consumed as food in the United States. The average pork and lard consumption per person in a typical year was 83 2 lb, whereas the consumption of other meats was 124.1 lb per person Pork is more successfully cured and stored than many other meats, and it is estimated that about 60% of the swine carcass is cured by various methods.

The United States has 24% of the world hog population. The 10 leading states, in order of number and value of hogs on farms, 1945–1954, are indicated in Table 1. The swine production industry

Table 1. Average number and value of hogs on farms, 1945–1954*

State	Number	Value
Iowa	11,116	$218,118
Illinois	5,941	117,642
Indiana	4,200	79,152
Missouri	3,717	61,486
Minnesota	3,474	66,921
Ohio	2,915	49,316
Nebraska	2,627	43,628
Wisconsin	1,717	31,859
Texas	1,623	16,060
Kentucky	1,356	19,609
United States	56,853	977,365

* USDA, *Agricultural Statistics*, 1956

exists as a part, varying from large to small, of a larger farm operation. Frequently, swine production is employed to supplement other farm operations so as to distribute labor requirements or to intensify the farm operation. However, the total number of hog producers is decreasing while those remaining in the business are increasing their volume. This trend is expected to continue.

Swine production embodies the application of the principles of breeding, management, nutrition, and marketing

Swine breeding. The swine breeding herd is either developed on a purebred breeding program or on a system of crossing the various breeds *See* BREEDING (ANIMAL). The breeds are rather arbitrarily divided into bacon- and meat-type breeds, with the latter the more popular The major breeds within the two classifications are listed below. A

Meat type		*Bacon type*
Berkshire	Minnesota No 1	Tamworth
Chester White	Minnesota No 2	Yorkshire
Duroc	Montana No. 1	
Hampshire	Spotted Poland China	

typical example of a meat-type breed is illustrated in Fig 1.

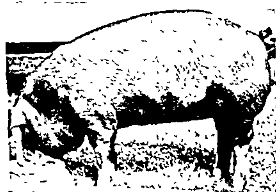


Fig 1. A typical example of a meat-type breed

Most of the hogs in the United States are produced in a crossbreeding program consisting of the mating of individuals of different breeds or inbred lines. Research has shown that crossbreeding increases the vigor and feeding qualities of the offspring; this effect is called heterosis or hybrid vigor. First-cross gilts or three-breed cross gilts, produced by mating first-cross gilts to a boar of a third breed, farrow and wean larger and heavier litters than purebreds. In practical swine operations a system of rotation breeding is recommended to realize the advantages of crossbreeding. Rotation breeding is accomplished by using male "seed" stock from three or four breeds in rotation and retaining the female stock sired by each boar.

Swine normally reach puberty between the ages of 4 and 5 months. Females exhibit sexual maturity at a slightly earlier age than males, but are not usually bred to farrow before they are 12–14 months old and weigh at least 225 lb. Boars are not used for breeding until they are at least 8 months old.

Estrus, or the time of sexual excitement in the sow, lasts from 48–72 hours and occurs every 18–24 days. Ovulation normally occurs during the latter part of the estrous period. It has been determined that female swine shed an average of 18 ova at each estrous period, although only 60–70% survive the gestation period of 114 days. The heavy prenatal mortality has not been explained.

Litter size, which averages 9.75 pigs, is influenced by a number of factors. The age of the female at mating is the most important factor; the number of pigs farrowed per litter increases up to the fifth litter, or the time the sow is about 3 years of age. Breeds differ, and within breeds the short, compact sows usually have smaller litters than the growthier individuals. About 5% of the pigs farrowed are stillborn.

Swine management. Items included as management are primarily directed at protecting the animals from adverse environmental factors and providing maximum opportunity for survival and growth. About 25% of the pigs farrowed fail to survive to a weaning age of 6–8 weeks. Crushing by the sow, chilling, and starvation are the principal causes of the heavy mortality rate. Farrowing stalls, or pens equipped with guard rails, and pig brooders are used to protect the newborn pigs (Fig. 2). During the first 2 weeks of the pig's life in a cool environment, it needs supplemental heat because its own heat regulatory mechanism is undeveloped. Clipping needle teeth, applying tincture of iodine to the navel, and administering sugar solution are also done to reduce mortality.

Since the pig has a very limited ability to dissipate body heat by sweating, fattening pigs or breeding stock must be protected during periods of high temperature. Shades, sanitary wallows, and water mists are used.

Swine are particularly susceptible to diseases and parasites. Cholera, erysipelas, dysentery, Traum's disease, leptospirosis, atrophic rhinitis, and transmissible gastroenteritis are some of the more important diseases (*see* ZOONOSES). In most cases, treatment of the disease is not well defined, and preventive or control measures are recommended. Swine are especially subject to infestation with ascarids, which are parasites primarily of the intestinal tract that impair metabolism and growth (*see* ASCARIS). The life cycle of the ascarid is complicated. Following ingestion of the embryonated ascarid eggs by the pig, the larvae burrow throughout the body. Eventually they find their way to the lungs and from there they migrate to the mouth. Then, following ingestion, the larvae develop into mature ascarids in the intestinal tract. Unembryonated eggs are passed in the feces. Hygromycin B, sodium fluoride, and piperazine compounds are effective anthelmintics.



Fig. 2 Farrowing stall with sow and litter.

The McLean-county system of swine sanitation has been proposed to control diseases and parasites. The scheme is composed of four steps: use of clean farrowing quarters, washing of sow prior to farrowing, sanitary removal of sow and litter to uninfested land, and retaining of pigs on uninfested land until at least 4 months of age.

Swine nutrition. Nutrition is an important aspect of swine production, because feed costs comprise about 80% of the total cost of production.

Feeding practices of today are scientifically developed to insure the maximum rate and efficiency of physiological performance. The performance noted in pigs fed rations which were characteristic at designated times in the history of the swine in

try is illustrated in Fig 3. Most of the progress resulted from an increased knowledge of the rient needs of swine.

Nutritionally, swine feeding is arbitrarily di- led into critical and noncritical phases. Gesta- n, lactation, and growth of the baby pig are con- dered critical periods, because nutrient needs are a peak The fattening period is noncritical Re- uirements of swine may be classified as water, nergy (carbohydrate and fat), protein (amino cids), minerals, and vitamins. It is essential that hese be in proper balance in order to promote op- imum performance.

absorp- nutri- y from

Most of the energy need of the pig is supplied by dietary carbohydrate and fat, although fat usually plays a minor role (*see* CARBOHYDRATE; FAT AND OIL, EDIBLE). In general, carbohydrates such as (except lactose),

may cause diarrhea in the weanling pig. but it is an excellent source of energy for the baby pig En- ergy requirements of swine are usually expressed as a daily need for total feed or digestible nutri- ents Feed intake increases as the pig matures be- use nutrient requirements for maintenance and ight gain are higher. For example, a 25-lb pig ill voluntarily consume about 2 0 lb of feed daily. hereas a 200-lb pig will consume about 7.5 lb

Protein is essential for maintenance growth, gestation, and lactation of swine Proteins consist of amino acids An essential amino acid is one that the pig cannot synthesize at a sufficiently rapid rate to permit normal performance The essential amino acids for the growing pig are arginine, his- tidine, isoleucine leucine, lysine methionine phenylalanine threonine, tryptophan, and valine (*see* AMINO ACIDS) In addition, cystine and tyro- sine can satisfy a part of the need for methionine and phenylalanine respectively The amino acids of most practical importance are lysine, trypto- phan and methionine which are inadequately sup- plied by corn Although there is an increasing em- phasis on amino acid needs of swine it is still customary to satisfy the need by providing ample quantities of protein.

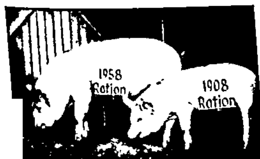


Fig 3 Progress in swine feeding (Courtesy of R J Meade, University of Minnesota)

The mineral and vitamin requirements of, and a typical ration for, the weanling pig are described in Table 2 (*see* MINERAL; VITAMIN). Corn lacks

Table 2. Nutrient needs and a typical diet of the weanling pig

Nutrient needs*		Typical ration	
Nutrient	Amount	Ingredient	Amount
Air-dry feed, lb	2 0	Yellow corn, lb	79 4
Total digestible nutrients, lb	1 5	Soybean oil meal, lb	18 0
Crude protein, %	18 0	Dicalcium pyro- phosphate, lb	1 0
Calcium, %	0 8	Ground limestone, lb	1 0
Phosphorus, %	0 6	Trace-mineralized salt, lb	0 5
Sodium chloride, %	0 5	Vitamin A and D supplement, lb	0 1
Copper, mg/lb	2 0	Riboflavin, mg	100 0
Iron, mg/lb	15 0	Calcium panto- thenate, mg	500 0
Iodine, mg/lb	0 1	Niacin, mg	800 0
Manganese, mg/lb	18 0	Choline chloride, mg	10,000 0
Carotene, mg/lb	0 75	Vitamin B_{12}, μg	700 0
Vitamin D, IU/lb	90 0	Antibiotic, g	1 0
Thiamine, mg/lb	0 5		
Riboflavin, mg/lb	1 2		
Niacin, mg/lb	8 0		
Pantothenic acid, mg/lb	5 0		
Pyridoxine, mg/lb	0 6		
Choline, mg/lb	400 0		
Vitamin B_{12}, μg/lb	7 0		

* National Research Council Committee on Animal Nutrition, *Nutrient Requirements for Swine*, 1953

many minerals and vitamins, hence practical diets normally require supplementation. Calcium, phos- phorus, sodium, and chlorine are normally added as ground limestone, steamed bone meal or dical- cium phosphate, and trace-mineralized salt. Trace minerals, such as iron, copper, manganese, iodine, and cobalt, are added as safety measures. Most

produced vitamin supplements.

Feed additives, such as antibiotic supplements, are becoming standard ingredients of swine rations (*see* ANTIBIOTIC) Antibiotic supplements contain- ing chlortetracycline (aureomycin), oxytetracy- cline (terramycin), or penicillin are added to growing pig rations to provide 5–10 milligrams of antibiotic per pound of ration. Antibiotics stimu- late the rate of gain, apparently by improving the general health of the pigs.

Swine rations are self-fed to growing-finishing pigs, since this conserves labor and results in max- imum feed intake Although most swine producers offer free access to shelled corn and supplement in self-feeders, there is a trend to feed mixed rations A supplement is a mixture of feeds which provides the nutrients that are lacking in corn Pigs have the ability to consume corn and supplement in amounts to constitute a balanced ration

With current feeding methods, pigs normally at- tain a market weight of 200–225 lb at approxi- mately 5½ months of age

Swine marketing. Of the swine slaughtered under Federal inspection, 87% consist of barrows and gilts and the remainder are sows, boars, and stags. The market hogs move from the farm to the slaughter plants through terminal public markets, dealers, concentration yards, truck buyers, auctions, and local cooperative associations, or by direct sale to the packer. The type of marketing varies considerably in different areas, although terminal public markets and direct packer purchase are the most popular channels.

Market grades of slaughter barrows and gilts have been developed according to a scheme which places a premium on the four major lean cuts, the ham, loin, picnic, and Boston butt. In addition, market grades penalize excess finish and lard Barrow and gilt carcasses are graded as U.S. No 1, U.S. No 2, U.S. No. 3, medium, and cull. Medium and cull grades are characterized by a low degree of finish.

Market price of hogs is largely determined by the prevailing commodity price level and by supply and demand. Keen competition from other meats has been particularly important in recent years. [D F.B.]

Bibliography: *See* AGRICULTURAL SCIFNCE (ANIMAL).

Switch, electric

A device which makes, breaks, or changes the course of an electric circuit. Basically it consists of two or more contacts mounted on an insulating structure and arranged to be moved into and out of contact with each other by a suitable operating mechanism *See* CONTACT, ELECTRIC.

The term switch usually is used to denote only those devices intended to function when the circuit is either deenergized or operating under normal load; as contrasted with circuit breakers, which also function when the circuit is carrying abnormal currents, such as short-circuit currents. *See* CIRCUIT BREAKER.

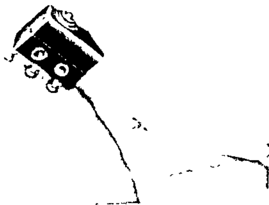


Fig. 1. Precision Microswitch. (*Minneapolis-Honeywell Regulator Co.*)

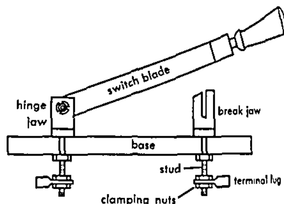


Fig. 2. Knife switch

Switches frequently are composed of many single circuit elements, known as poles, all operated simultaneously or in a predetermined sequence by the same mechanism Switches used in complex machines, such as computers, may have a large number of poles. Switches used in power circuits usually have from one to four poles, depending on the kind of circuit. Switches are often typed by the number of poles, such as single-pole or double-pole It is also common to express the number of possible switch positions per pole, such as a single-throw or double-throw switch.

Types of switches. Switches are made in many forms, ranging from Microswitches, used in low voltage circuits carrying small currents in such equipment as telephones and business machines, to disconnecting switches in power transmission circuits carrying thousands of amperes at about 500,000 volts. They are classified in hundreds of categories according to their use or construction.

Most familiar are the wall switches used in homes and offices for turning lights and appliances on and off, canopy switches in portable lamps, dial and pushbutton switches on electric ranges and automatic washing machines and dishwashers, and service entrance switches used to cut off power from buildings in emergencies.

Knife switch. The oldest and most symbolic form of switch, the knife switch, consists of a metal blade hinged to a stationary jaw at one end and contacting a similar jaw at the other end (Fig 2) A vast array of switches, from the largest high-voltage disconnecting switches more than 30 ft high and 20 ft long to small low-voltage switches the size of a matchbook, are refined adaptations of this simple form.

Leaf-spring switch. In this type of switch, parallel strips of spring metal are sandwiched between blocks of insulation at one end and pushed into or out of contact at the other end by a suitable operating mechanism These are extensively used in communication equipment and other light machines.

Sliding-contact switch. Frequently used where several different circuits are to be switched into different patterns, these take the form of a dial or drum with metal segments making contact with contact fingers sliding over the dial or drum surface Examples of these are station-selector

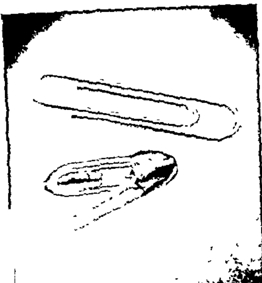


Fig 3 Miniature mercury switch element may be used with a variety of actuating mechanisms to cause shifting (Minneapolis-Honeywell Regulator Co)

switches in some television receivers and tap-changing switches in power transformers.

Butt-contact switch. This is another widely used form, in which a short movable bar of metal bridges the gap between two fixed conductors and is moved into and out of contact by a variety of *mechanisms*. Cam operated master switches used extensively on machine tools and control switches on power switchboards are [illegible]

pool of mercury (Fig 3), providing silent operation and isolation from explosive and corrosive gases

Arc extinction. Switches known as load-break switches used in power applications are often required to *interrupt* considerable current To do this safely they must be equipped with special devices for quickly extinguishing the arc that occurs when the circuit is broken. Known as arc quenchers, these usually take the form of chambers of ceramic or other heat-resisting materials surrounding the arc. Sometimes a blowout coil is used to drive the arc rapidly along diverging *contact surfaces* reaching into the arcing chamber, lengthening the arc and hastening its extinction (*see* BLOWOUT COIL).

Successful arc extinction in any switch is expedited by rapid motion of the switch contacts. To accomplish this, a variety of mechanisms is used, the most common of which is the over-center toggle mechanism found in the ordinary wall switch In this device, the first movement of the handle stores energy in a spring and further movement releases the spring to snap the switch open or closed.

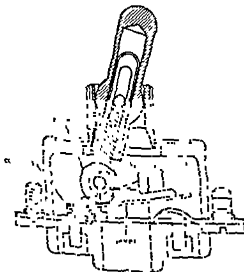


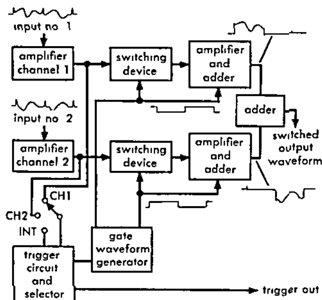
Fig 4 Toggle-switch mechanism

Special switch applications. Because of their size and *frequently inaccessible* location, high-*voltage* switches and low-voltage switches of high current capacity are often operated by power-driven mechanisms controlled from a *remote point* by small *control* switches A large category of *switches known as* auxiliary, limit, and position switches are arranged *for attachment* to all manner of *machines to be* operated by the motion of *their parts rather* than by hand. [W.N.G.]

Switch, electronic

An electronic device to *which two input* waveforms can be applied *and* which delivers at a pair of output terminals a signal that is alternately a replica of each of the input signals. The transmission gate performs the basic switching *function of* the *electronic switch*, and sometimes such a gate circuit is itself defined as an electronic switch (*see* GATE CIRCUIT) However, *the electronic* switch is usually considered to be *a separate* instrument, to which, periodically recurrent input signals, of arbitrary waveform but with synchronously *related* periods, are connected The output is *switched* between the two waveforms at a rate that is synchronous *with the* period of the input *waveforms*. One of the input signals often provides *synchronizing* information to supply *periodic* trigger pulses, which are then applied *to* the gate signal *generators* within the switch

A frequent use of the electronic switch is to provide *means* for displaying two *time* related signals on the screen of a cathode-ray oscilloscope without requiring two independent deflection systems *within the* cathode-ray tube For this application, the time-based circuit of the display device is synchronized with the repetition *rate* of one of the two input signals, and the internal transmission gate of the switch alternately switches the signals to the output. At the same *time* a dc component, or pedestal is added to each of the two signals so that they will appear at distinct levels on the screen of the display device.



Block diagram of elements of electronic switch

A block diagram of an electronic switch is shown in the illustration. The two input waveforms are applied to amplifier channels. The amplified waveforms are then applied to switching devices and also to a trigger circuit, which generates trigger pulses that can be synchronized either internally or with one of the input signals. A gate waveform generator, usually a bistable multivibrator, is used to actuate a gating waveform to operate the switching devices alternately.

A controllable portion of the gating pulse may be combined with each switched channel to provide an amplitude separation of the two switched signals where separation is necessary for such applications as the cathode-ray tube display

The trigger circuit might supply a trigger to the time-base generator of an oscilloscope, or when internal triggers are generated, it might be used to control the repetition period of the phenomena to be displayed. *See* WAVE-SHAPING CIRCUITS [G.M C.]

Switching circuit

A constituent electric circuit of switching or digital data-processing systems. Well-known examples of such systems are digital computers, dial telephone systems, automatic accounting and inventory systems. In these and other switching systems the component circuit units receive, store, and manipulate information in coded (digital) form to accomplish the specified objectives of the system. *See* SWITCHING SYSTEMS (COMMUNICATIONS); SWITCHING THEORY.

Physically, switching circuits consist of conduct- ... discrete-valued electrical ... circuit ... witches and relays in which manual or electromagnetic actuation opens and closes electric contacts; vacuum and gas-filled electronic tubes, semiconductor rectifiers and transistors, which do or do not conduct current; and magnetic structures, which can be saturated in either one of two directions.

The electrical conditions controlling these switching circuit devices are also generally two-valued or binary, such as open versus closed path, full voltage versus no voltage, large current versus small current, and high resistance versus low resistance. Such two-valued electrical conditions, as applied to the input of a switching circuit, represent either (1) a combination of events or situations which exist or do not exist, (2) a sequence of events or situations which occur in a certain order, or (3) both combinations and sequences of events or situations. The switching circuit responds to such inputs by delivering at its output, also in two-valued terms, new information which is functionally related to the input information.

The two fundamental characteristics of switching circuits are logic and memory. A switching circuit embodies such logical relationships as output X is to exist only if inputs A and B occur simultaneously; and output Y is to exist if either input A or input B occurs. The factor of memory, in turn, enables a switching circuit to hold or retain a given state after the condition that produced the state has passed.

Basic combinational circuits. A combinational switching circuit is one in which a particular set of input conditions always establishes the same output, irrespective of the past history of the circuit. An example of a simple combinational circuit is the problem of controlling the entrance-hall light of a residence by three up-down wall switches located in three different rooms; that is, any one of the three wall switches must be able to turn the hall light either on or off Analysis of this problem shows that the circuit must meet the following simple requirements If any one or all three wall switches are down, the hall lamp must light; if any one or all three switches are up, the lamp must be dark. An obvious (but not the most efficient) circuit meeting these requirements is shown in Fig 1

In this problem the circuit inputs are, of course, the manual switch settings, and the circuit output is the control of the light.

In electronic switching circuits, so-called gates are used to perform logical functions equivalent to these series-parallel networks of switch contacts

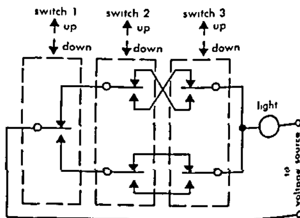
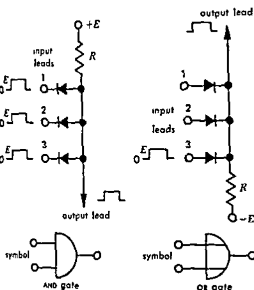


Fig 1 Elementary combinational switching circuit



ig 2 Typical switching gates using crystal diodes

n this sense, an electronic gate is an elementary ombinational circuit Gates do not function by hysical rearrangement of interconnecting paths, is do switch or relay contacts Instead, they funcion by control of voltage or current levels at their utput.

The most commonly encountered gates are the AND, and the OR gates. The AND gate produces an output only if all its inputs are concurrently present, an OR gate produces an output if any one or any combination of its inputs is present Figure 2 shows both an AND gate and an OR gate, using rectifier or diode elements

In the AND gate the rectifiers are so oriented that current from a positive voltage source E passes through the relatively large resistance R and then through the low forward resistance of any one of the rectifiers to ground in the circuits controlling the gate Thus, in the inactive state of the gate the output lead is at or near ground potential. If all three input leads of this gate concurrently receive a positive voltage pulse of magnitude E, the rectifiers approach open circuit and the output lead will be raised from near ground to a positive potential for the duration of the input pulse. In other words, input leads 1 AND 2 AND 3 must all receive the positive pulse to obtain the positive output voltage.

In the OR gate the rectifiers are reversed so that current flows from ground in the input circuits through the low forward resistance of any rectifier and then through the relatively large resistance R to the negative voltage source $-E$. Thus, in the inactive state of the gate the output lead is at or near ground potential If, however, a relatively high positive voltage pulse is applied to input leads 1 OR 2 OR 3, the remaining two diodes are cut off and the output is raised to a positive potential for the duration of the input pulse.

Gates may, of course, be constructed with other electronic devices, such as tubes, transistors, and magnetic cores.

Basic sequential circuits. A sequential switching circuit is one whose output depends not only upon the present state of its input, but also on what its input conditions have been in the past. Sequential circuits, therefore, require memory elements.

By way of illustration, consider the following simple sequential circuit problem. When a tele-

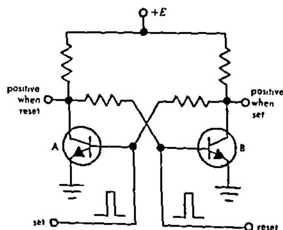


Fig 4 A transistor switching memory element (flip-flop).

phone customer l...

...ing. After the operator has satisfied the customer's request for a connection, she withdraws The light, however, should not relight now, even though the conditions existing at this time are seemingly identical with those at the start; that is, the customer has his handset lifted and no operator is on the line A sequential relay circuit meeting these simple requirements is shown in Fig. 3. In this circuit, when the handset is lifted, the handset off-hook switch connects a ground input to relay A which operates and lights the switchboard lamp When the operator answers, another ground input operates relay B, and this relay puts out the light. A holding circuit on relay B keeps

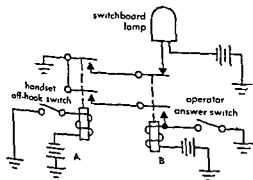


Fig 3 Elementary sequential relay switching circuit.

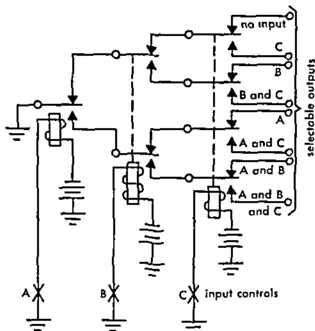


Fig 5. Simple relay selecting circuit.

relay B operated until the handset off-hook switch is again opened and relay A is deenergized. Relay B "remembers" that the operator has answered and prevents the relighting of the lamp when the operator withdraws. It is, therefore, the memory element of the circuit.

A typical electronic memory element used in sequential circuits is a simple circuit called a flip-flop. A flip-flop consists of two amplifiers connected so that the output of one amplifier is the input of the other. A voltage pulse will set the flip-flop into one of two states, and that state remains until another voltage pulse resets, or returns the flip-flop to its original state. It can therefore be used to remember that an event has taken place

Figure 4 is an *n-p-n* transistor flip-flop. When set, transistor A is conducting and transistor B is cut off When reset, transistor B is conducting and transistor A is cut off. A positive output voltage with respect to ground may be obtained from either transistor to indicate the condition of the flip-flop

Relays, flip-flops, and similar memory elements provide static, or fixed, memory; they hold the stored information indefinitely, or until they are told to "forget." In contrast, a delay-line provides transient memory. A delay line has the property that an electrical signal applied to its input is delayed on its way to the output.

Functional switching circuits. Even in large and complex switching systems the majority of circuit requirements can be met by a relatively small number of types of circuits, each of which performs one or a limited number of somewhat distinct functions. These functional circuits are the basic building blocks of a switching system.

Selecting circuits. A selecting circuit receives the identity (called the address) of a particular item and selects that item from among a number of similar ones. The selectable items are often represented by terminals or leads. Selection usually involves marking the specified terminal or lead by applying to it some electrical condition, such as a voltage or current pulse, or a steady-state dc signal. By means of this electrical condition, the selected circuit is alerted, seized, or controlled.

Figure 5 is a simple relay selecting circuit This circuit uses three relays to select one of eight outputs according to the combinations in which the three relays are operated or not operated The input is ground or no ground on control leads A, B, C.

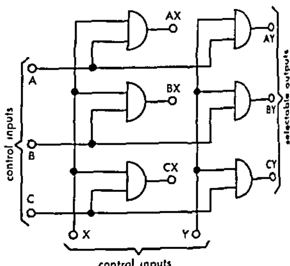


Fig. 6. Matrix selecting circuit using AND gates

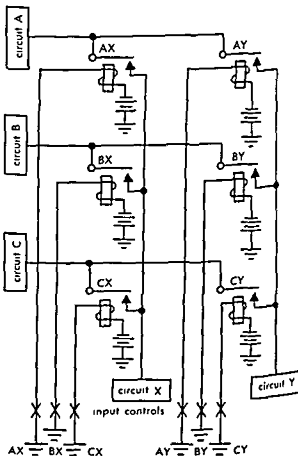


Fig 7. Simple relay connecting circuit.

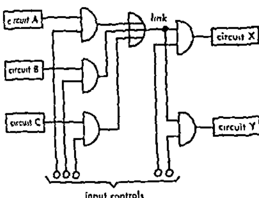


Fig 8 Connecting circuit using AND and OR gates.

in various combinations (the address) The output of the circuit is ground appearing on the single selected output lead

An electronic selecting circuit using AND gates is the matrix type shown in Fig 6 In this type of circuit an input signal appears on one of the horizontal input leads and concurrently on one of the vertical input leads. The selected output is at the crosspoint of these two leads.

Connecting circuits A switching system is an aggregate of functional circuit units, some of which must sometimes be directly coupled to each other to interchange information. Such a system needs, [illegible]refore, connecting circuits which establish the [illegible]cuit associations dictated by the momentary [illegible]ds of the system.

Figure 7 is a simple relay circuit to illustrate the [illegible]nciple of connectors. Any one of the three cir[illegible]its A, B, or C can be connected over a single [illegible]

[illegible]ctual control circuit that determines which cir[illegible]it association is needed and insures that only one [illegible]lay is operated at a time in any row or any col[illegible]nn The connector relays may, of course, carry more than one interconnecting lead, and the number of interconnectable circuits could be fewer or many more

Figure 8 shows a simple electronic connecting circuit using AND and OR gates. In this arrangement a communication path is provided over a single link from any one of the three functional circuits A, B, C, to either the X or Y circuit by an external control circuit activating the appropriate pair of AND gates To provide a multilead link, or to provide for other simultaneous interconnections, additional AND gates would, of course, be required. The OR gate maintains separation of the inputs at the common junction point.

Lockout circuits. In switching systems, situations often arise where several similar circuit units are ready at the same instant to request collaboration with another type of functional circuit. Mutual interference among the requesting circuits is prevented by the lockout circuit. In response to concurrent inputs from a number of external circuits, a lockout circuit provides an output indication corresponding to one, and only one, of these circuits at any time.

Figure 9 illustrates a basic relay lockout circuit. The external circuits signify their requests to be allowed to proceed by grounding their respective control leads designated C. The output of the lockout circuit is ground appearing on a single lead designated B, associated with the particular external circuit whose request has been granted. The characteristics that enable this circuit to perform its function, are (1) the output ground goes through a contact network chained from left to right; this ground can appear only on the output lead of the lowest numbered operated relay which represents the winning external circuit; (2) the voltage source or battery on which the relays operate, in turn goes through another contact network chained in such a manner that once any relay operates, from then on only higher numbered relays are permitted to operate; (3) an operated relay

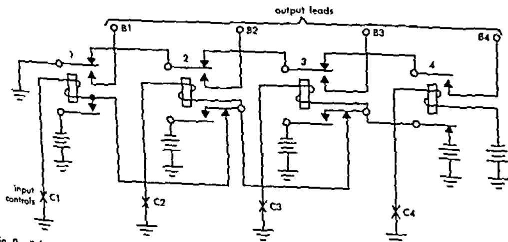


Fig 9. Relay lockout circuit

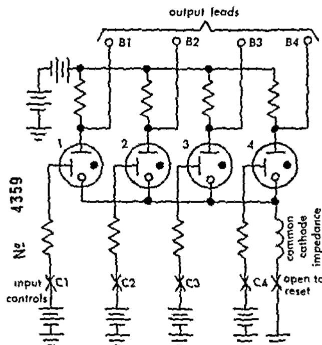


Fig 10. Lockout circuit using cold-cathode gas tubes

stays operated on battery through its own closed contact, until the external circuit removes the control ground as an indication that it has been satisfied.

Figure 10 shows a typical electronic lockout circuit using cold-cathode gas-filled tubes. The external circuits furnish positive potential on the input leads to the control gaps of the tubes as indications of service requests The operation of the circuit is based on the dynamic negative-resistance characteristics of gas tubes. If such tubes are provided with a common impedance in their conduction paths (the cathode impedance in this circuit), simultaneous input signals will result in the ionization of only one tube Once the control gap of a tube is ionized, conduction current starts flowing in its main gap and this current through the common impedance instantaneously reduces the voltage across all the other tubes below the value needed to ionize them. This reduced voltage is, however adequate to keep the single ionized tube in the conducting state until its conduction path is opened The identity of the particular ionized tube is derived from the anode resistance individual to each tube; the output lead whose potential has been lowered by this resistance represents the circuit whose request has been granted.

Lockout circuits are sometimes referred to as hunting or finding circuits. Irrespective of name, the problem in all applications of lockout circuits is that of concurrently competing circuits, among which one has to be picked for some action

Translating circuits. Switching systems process information in coded form; the information they receive and manipulate is generally in the form of numbers.

Numerical codes are many and varied, each with its own characteristics and more or less distinct advantages for different switching circuit situations Therefore, one of the common functional circuits in switching systems is the translating circuit, which translates information received in one code into the same information expressed in another code These translating circuits are combinational circuits; a given input signal combination representing a code to be translated always produces the same output signals, which represent the desired code.

Figure 11 is an elementary relay translating circuit In this circuit the input code is biquinary (ground on one of the input leads 1 to 5 being the quinary or five-valued part, and ground on lead A or B being the binary or two-valued part). The output of the circuit in turn is decimal; in response to a biquinary input, a ground appears on one of the 10 output leads.

Figure 12 is an example of a magnetic-core translating circuit that translates from binary code (1, 2, 4) to a one out of eight code (0, 1, 2, . . . , 6, 7). The circuit has three flip-flops which are set

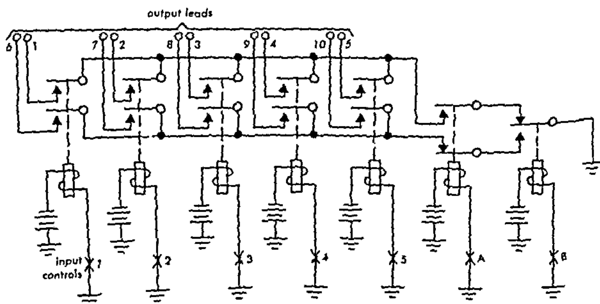
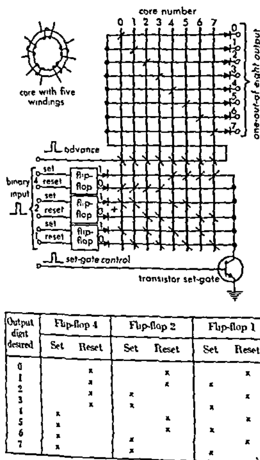


Fig. 11. Simple code translating circuit using relays.



Output digit desired	Flip-flop 4		Flip-flop 2		Flip-flop 1	
	Set	Reset	Set	Reset	Set	Reset
0		x		x		x
1		x		x	x	
2		x	x			x
3		x	x		x	
4	x			x		x
5	x			x	x	
6	x		x			x
7	x		x		x	

Fig 12. Code register and *translating circuit* using magnetic cores

or reset (not set) according to the binary input code combination. The translating elements are eight magnetic cores, each with five windings, and are represented in Fig 12 by a heavy vertical line Each short, slanting line segment represents a separate winding on a core These short slanting lines also symbolize a *mirror action*; an input current pulse coming from a flip flop sets a core if it is reflected upward by the mirror and prevents setting or resets the core if reflected downward Once set, the subsequent resetting of a core induces a current which flows upward in the vertical line (in a direction opposite to the resetting current) and is reflected left [illegible]

[illegible]

[illegible] of a positive voltage pulse to each of the three flip-flops either on its set or on its reset input lead, according to the table. (Note that by adding the numerical designations of those flip-flops which are set in a particular combination, the value of the output digit is determined)

While the flip-flops are being set, their output current is prevented from flowing into the core windings by the transistor set gate which is normally nonconducting Shortly after the binary input combination is recorded in the flip-flops, this set-gate is pulsed for a moment into its conducting state. During this moment, output current will flow from each flip-flop either in its ONE output lead (if the flip-flop has been set) or in its ZERO output lead (if the flip-flop has been reset). As Fig. 12 shows, the output current of flip-flop 4 is always used *to set the cores; that is, the current in the* ZERO output lead of this flip-flop is used to magnetize the first four cores in the set direction, or the current in its ONE output lead is used to magnetize the last four cores in the set direction. In contrast, the output currents from flip-flops 2 and 1 are always used to magnetize the cores in the opposite or reset direction. Initially all cores are in the reset condition, and cores that receive both set and reset currents simultaneously will not change this initial condition. An analysis of Fig 12 will therefore show that, for any desired digit, one and only one of the eight cores will be set by the flip-flops in combination. For instance, if output 3 is desired, the current from flip-flop 4 tends to set cores 0, 1, 2, and 3, but cores 0, 1, and 2 are prevented from being set by the output current from either or both flip-flops 2 and 1. When the translated code is needed, the current pulse on the advance lead resets the single previously set core, and consequently an induced output current pulse appears on the appropriate output lead (The rectifiers in the input and output portions of the circuit prevent unwanted reverse current.)

Register circuits Information received by a switching system is not always used immediately. It must be stored in register circuits for future use.

In a register circuit the coded information to be stored is applied as input, is retained by memory elements of the circuit and, when needed, the registered information is taken as output in the same code or in a different code Figure 12 embodies a register function as well as a translating function Register circuits are devised with a great variety of memory elements and have capacities to store from a few to millions of [illegible]

[illegible] to shift its stored digital information internally to positions representing higher or lower numerical values in the code employed. For example, in decimal code registration a digit may be shifted from the units to the tens position. An obvious use of such registers is in digital computers when, for example, partial multiplication products have to be lined up for addition.

Counting circuits One of the most frequently encountered circuits in switching systems is the counting circuit whose function, in general, is to detect and count repeated current or voltage pulses which represent incoming [illegible]

[illegible] *Logical Design of Digital Computers*, 1958

Switching systems (communications)

The assemblies of switching and control devices provided so that any station in a communications system may be connected as desired with any other station. In telephone practice the switching centers are known as central offices.

Numbering plan. The numbering plan must uniquely identify each telephone main station so that calls may be directed to it. The minimum requirement is for an adequate number of digits or *characters in each number to provide for each* main station in the dialing area.

A telephone central office customarily has capacity for serving 10,000 main telephones. Within a central office a series of 4-digit numbers, 0000 to 9999, is used to designate each main station. Some of these may be unused at a given time

When a town requires more than 10,000 main telephones, more than one central office is provided. Each office is then given a separate designation. One or two numerical digits, word names, and word names plus a numerical digit have been used. These central office designations are known as central office codes. In 1952 the industry in North America agreed to standardize and started conversion to the 2-letter, 5-numeral (2L–5N) numbering plan. The two letters are the first two in a word name, for example, ADams 2-9970 (equivalent on the dial to 232-9970). The two letters plus the first numeral (AD2) constitute the central office code. The switching equipment or the operator needs to consider only the central office code in order to direct a call to the proper office. When the called office is reached, the main telephone wanted is determined from the last four numerals.

A 7-digit numbering plan does not have adequate capacity for all the telephones in North America. Hence, a geographical area, such as a state or a Canadian province, is selected as a numbering plan area (NPA) within which there are no duplications of 2L–5N numbers. Although there are 640 possible central office codes in the 2L–5N numbers, only about 500 are satisfactory for general use. Populous states have more than 500 central offices and so are divided into two or more numbering plan areas

Each numbering plan area is given a 3-digit NPA *code, the middle digit of which* is always a 1 or 0 Examples are 703 for the state of Virginia and 415 for the portion of California that includes San Francisco, as shown in Fig. 1. In dialing a number of a subscriber outside the local or home numbering plan area, the area code should be dialed ahead of the 2L–5N number. For example, a subscriber or operator in Asbury Park, New Jersey wishing to dial GArfield 1-9000 in San Francisco would first dial the area code 415 followed by GA 1-9000 From any other numbering plan area, the dialing would be identical to reach the GArfield 1-9000 number, except from within the 415 area, where only the seven digits of the telephone number need to be dialed. With this plan the equipment uses first the NPA code to determine which NPA area is desired, second the central office code to select the office in that area, and third the main telephone number to determine the particular telephone being called. This service is known as direct distance dialing.

In addition to central office and area codes, code numbers are used for special services, such as 0 to reach the operator, 411 or 113 for information, and 611 or 114 for the repair desk.

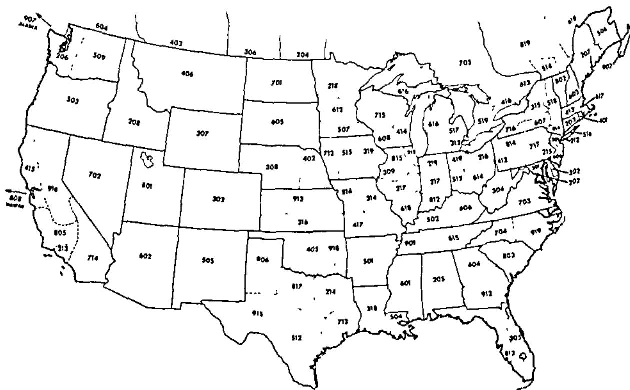
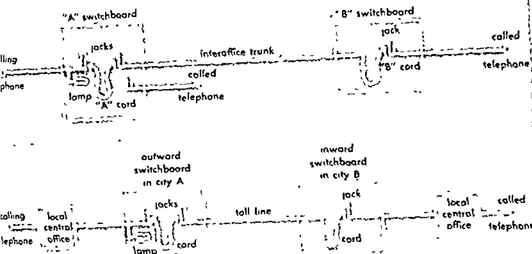


Fig 1. Numbering-plan areas with codes



2. Manual switching systems (a) Local (b) Toll.

Manual switching system. In a manual system, ...ephone connections between calling and called ...rties are established manually by operators in ...

... between cities or towns.

Figure 2*a* shows a typical local telephone connection The calling telephone connects to the A (outgoing) switchboard in a jack and a lamp, which lights when the calling telephone handset is removed. The operator answers by inserting the ...

... jack of the called party if that party is also connected to the same A switchboard If the *called* party is served by another office, the A operator routes the call over an interoffice trunk to the B (incoming) switchboard in that office and passes the order verbally to the B *operator*, who inserts the plug of the single-ended B cord in the jack of the called telephone.

Local manual switching systems are rapidly being replaced by dial systems; in 1959, 95% of United States telephones were the dial type.

Figure 2*b* shows a typical toll *connection* between manual toll switchboards in two cities. The *calling telephone* is extended through the local office (either manual or *dial*) *to the toll* office, where it is answered by a *toll* operator at an outward switchboard known as the combined recording completing switchboard Here a record of the call and its *duration* is *made* for billing purposes. In completing the call, *the* operator uses a toll line to the distant city and passes the called telephone *number verbally to* the inward operator, who in turn completes the call *through* the local office (either manual or dial) to the called telephone.

Direct toll lines between cities A and B *may not always exist.* In this case, the toll line is routed

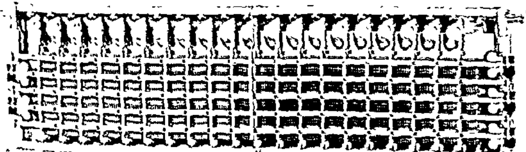


Fig 3 Crossbar switch

through an intermediate location where a toll tandem switchboard acts as a distributing point.

The toll system is also being mechanized, the inward and tandem equipment being replaced by dial switching. The outward operator is then the only operator on the connection.

Crossbar system. Two outstanding features of the crossbar switching systems as used in the United States are (1) the crossbar switch itself and (2) the marker system of controlling the crossbar switches in the setting up of the talking connections from calling to called parties.

Figure 3 shows a crossbar switch. It consists basically of (1) 20 vertical circuit paths, (2) 10 horizontal circuit paths and (3) a mechanism permitting any vertical path to be connected to any horizontal path. The switch, therefore, has 200 crosspoints. The contacts of the horizontal and vertical paths are the relay type. There are five crossbars, from which the switch gets its name. Each crossbar is located between two rows of horizontal paths and can be rotated through a small arc in either direction by means of two electromagnets and a butterfly-shaped armature. Rotating the bar to the upward or downward position causes the selecting fingers to move to the upper or lower contacts. The operation of a vertical-hold magnet, located at the bottom of each vertical path, then closes the crosspoint. By operating one of 10 selecting magnets and one of 20 vertical-hold magnets, it is possible to close any particular crosspoint. A crossbar switch is also made with 100 crosspoints instead of 200. Each crosspoint may have 3, 4, 5, or 6 pairs of contacts.

Figure 4 shows the fundamental method of using the crossbar switch. Here 20 lines are connected to the 20 vertical paths and 10 trunks to the horizontals. Any one of the 20 lines can be connected to any of the 10 trunks by operating the proper crosspoint. Any number of lines may be given access to the same 10 trunks by adding other switches and extending the trunk lines to the added switches. Unlike other switches, which can handle only one call at a time, the crossbar switch may have as many as 10 calls through it at one time. However, setting up of connections must be done one at a time.

A method of using the crossbar switch to obtain greater trunking access is shown in Fig. 5. Here 10 switches are used as primary switches and another

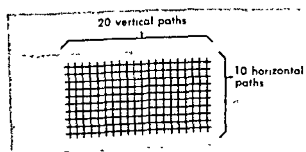


Fig. 4. Arrangement of crosspoints on crossbar switch.

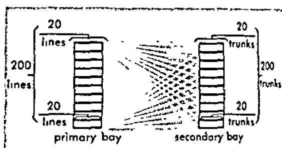


Fig. 5. Primary-secondary trunking arrangement

10 as secondary switches. The 10 horizontals of a particular switch on the primary bay are wired to one horizontal on each of the 10 switches in the secondary bay. The other nine switches are similarly wired to other horizontals on the secondary bay, providing an arrangement in which any one of 200 vertical paths on the primary bay may be connected to any of the 200 verticals on the secondary bay by operating the proper primary and secondary crosspoints. Several variations of this basic scheme are possible, giving different trunking arrangements.

Connections in a crossbar system are controlled by an assembly of many relays known as a marker. The crossbar switch, the multicontact relays, and most other relays are constructed of fast-operating short-movement, relay-type structures. With fast operation, the time required to set up a connection is small and consequently a small number of markers and other common control equipment is sufficient to handle the calls at high calling rates.

The American crossbar system is unlike the previous dial systems (the panel type or the step-by-step type) in which the switches are set up one after another. When a call is originated in a crossbar office, the location of the calling line in the switches is marked. The location of the called line or outgoing trunk to another office is also marked. The marker then selects an idle talking channel through the crossbar switches to interconnect the marked points and causes all contacts in this channel to be closed simultaneously. The channel is held busy for the duration of the call.

Also, unlike previous dial systems, the marker of the crossbar systems is arranged to look at alternate routes to the called office in case all trunks of the first-choice route are busy. The crossbar circuits are designed so that the marker can detect certain trouble conditions in the office and make a second trial over other circuits to complete the calls. A record of the trouble and its probable location is made for the maintenance people.

No. 5 crossbar switching system. The latest form of crossbar equipment for local central office use is the No. 5 crossbar system. Although local switching is its primary use, it can be adapted for switching toll lines and tandem trunks where there are not enough such lines or trunks to justify separate toll or tandem switching systems.

The system employs two types of switching frames, (1) the line link (LL ames on which the

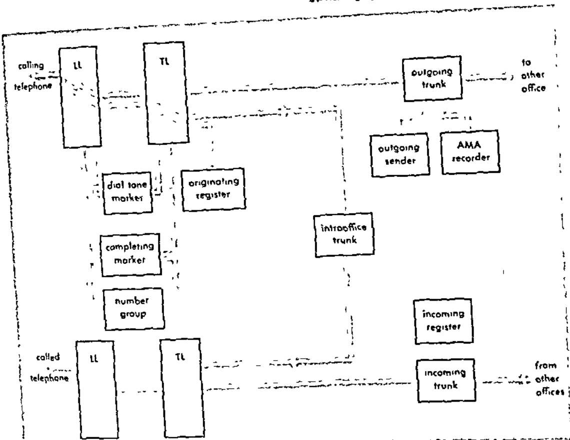


Fig 6. Diagram of trunking in No 5 crossbar office

telephone lines appear and (2) the trunk link (TL) frames on which the trunks appear Each frame consists of a primary-secondary switching network A maximum of 60 LL and 30 TL frames constitutes a unit, or marker group, of equipment Each LL frame is interconnected with every TL frame by a network of tie lines, or junctors Each LL frame has a basic capacity for 290 telephone lines. By adding supplementary crossbar switches to this basic frame, its capacity may be increased in 50 line increments to [illegible]

calling telephone is lifted to initiate a call, the dial tone marker causes the calling telephone to be connected through the LL and TL frames to an idle originating register The register then places dial tone on the line as an indication for the customer to begin dialing If the called number is in the same office as the calling telephone, a completing marker is chosen to establish the connection, and the called number together with the LL frame location of the calling telephone, are transferred into it The completing marker consults a number group to find the LL frame location of the called number. An idle intraoffice trunk and channels through the TL and LL frames are chosen to interconnect these two locations The corresponding crosspoints are closed and the called telephone is rung The connection to the originating register is then dropped.

If the call is to a telephone in another central office building, the completing marker selects an outgoing trunk (rather than an intraoffice trunk) to the called office. The called number is transferred from the completing marker into the outgoing sender The outgoing sender pulses forward the called number to the terminating point.

A call originating in some other office for a number in the crossbar office reaches the office over an incoming trunk to which an incoming register is temporarily connected. The called number is pulsed from the originating office into the register. The register associates itself with a completing marker which, with the help of the number group, selects and closes the channels through the TL and LL frames to the called telephone.

The No 5 crossbar equipment has been designed to handle 2L–5N dialing for dialing within the local numbering plan area It is also arranged for direct distance dialing (DDD) service using the national numbering plan area codes.

An important feature of the system is the method of recording the calls for billing purposes This method is known as automatic message accounting (AMA) and consists of perforating the data on a broad paper tape On calls to the local area where a message-unit basis of charging applies, only the calling telephone and the number of message units

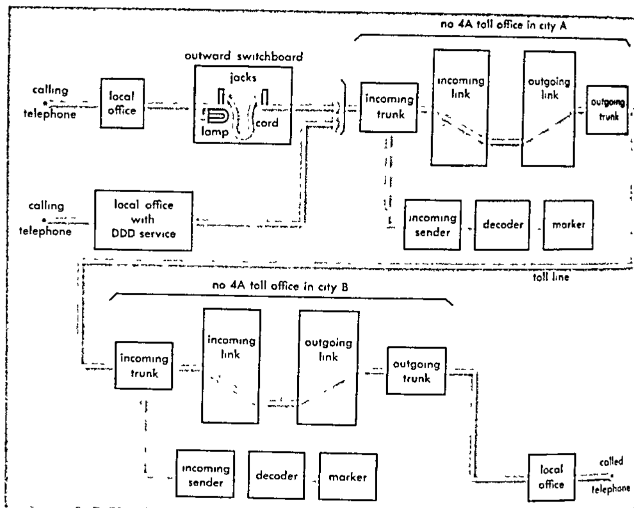


Fig. 7 Diagram of trunking in No. 4A toll crossbar office

is perforated on the tape For toll calls, the calling and called telephones, answering and disconnect times (from which the length of conversation is computed) and other pertinent data are perforated on the tape. Automatic machines later decipher the perforations and transcribe the data to printed form.

Another important feature of the No 5 crossbar system is its ability to interconnect with all types of switching systems. It is arranged to send out and to receive multifrequency pulses between No 5 crossbar offices, dial pulses to and from offices using step-by-step equipment, revertive pulses for panel and No 1 crossbar offices and call indicator pulses to manual offices.

No. 4A toll crossbar switching system. This system is commonly used in the completion of toll calls between distant cities. It is a means for mechanizing the handling of toll calls It is designed to handle the traffic in the largest cities

Figure 7 shows a typical arrangement of the switching system with a No. 4A toll crossbar office in each of two cities. It also shows at the originating point two methods of placing calls from the calling telephone: (1) that in which the call is routed through the local office to an operator who dials the call at the outward switchboard and (2) that in which the calling telephone is served by a local office with DDD service. In (2) the calling customer dials the call, and it is completed without the aid of operators.

The 4A crossbar system has two main switching frames. Incoming trunks from originating points appear on the incoming link frame, and the outgoing trunks leave from the outgoing link frames Each frame consists of a primary-secondary switching network.

A unit or marker group of equipment may have a maximum of 40 frames of each type, giving it capacity for terminating 8000 incoming trunks and 12,000 outgoing trunks. The number of terminations could be doubled by adding a second set of markers There is full flexibility for any incoming trunk to be connected to any outgoing trunk.

A call arriving at the No. 4 crossbar office in city A, either from an operator or from a customer in an office with DDD service, appears at the incoming trunk and is connected to an incoming sender, into which the called telephone number (2L–5N or national area code plus 2L–5N) is pulsed The decoder determines the routing of the call from this number and connects to an idle marker, passing the routing information to it. The marker closes through the channel to interconnect the incoming trunk with the outgoing trunk, which in Fig. 7 is to another 4A crossbar office in city B The call arriving at city B is handled in the same manner as at city A, the 2L–5N number being

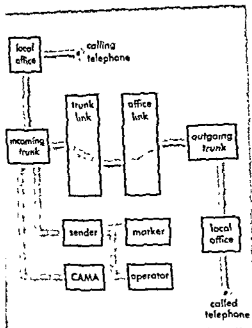


Fig 8 Diagram of crossbar tandem switching system

pulsed from the sender in A to the sender in B The decoder in B determines the routing and completes the connection to the local office which in turn sets up the connection to the called telephone

An important feature of the No 4A toll crossbar system is its ability to pick an alternate route if the first-choice route between cities A and B has no idle toll lines For example, a route from city A to city C to city B may be selected if no direct line to city B is available Several such alternate routes may be lool...
everal sw...
iating toll...
good servi...
route is busy, there is a good possibility that a toll line in some other route may be idle

The toll switching system in cities B and C need not be of the No. 4A crossbar type They frequently are of the step-by-step type, the crossbar tandem type or, for small offices, the No. 5 crossbar type. The No. 4A equipment is capable of sending ...

as required to operate the various switches

The No 4A system is designed for 4-wire switching; the voice is carried over one pair of wires in one direction, and the other pair of wires carries the voice in the opposite direction.

Crossbar tandem switching system. This system is used for switching of local interoffice trunks, or of intertoll trunks in a toll network.

Like the other crossbar systems, it has two types of switching frames, trunk link frames and office link frames, of which there may be 20 each in a marker group. This system has capacity for a maximum of 3200 incoming and 4000 outgoing trunks. Supplementary frames may increase these to 6400 and 6000, respectively. Its capacity is smaller than that of the No. 4A toll crossbar and accordingly it is used where the larger capacity of the No. 4A system is not needed.

Figure 8 shows the trunking plan, which is similar to those of the other crossbar systems. The functions of the sender and marker are also the same

Centralized automatic message accounting (CAMA) has been applied to the tandem equipment. The CAMA equipment is similar to that used in No 5 crossbar offices. Instead of the calling telephone being identified automatically as in the No 5 crossbar, an operator is momentarily connected She asks the calling party for his number and pulses it into the sender. The system is also adaptable to the installation of automatic number identification in the local offices In this event the operator at tandem is not needed; the calling tele-

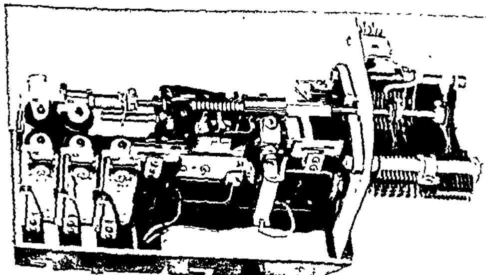


Fig 9 Step-by-step switch

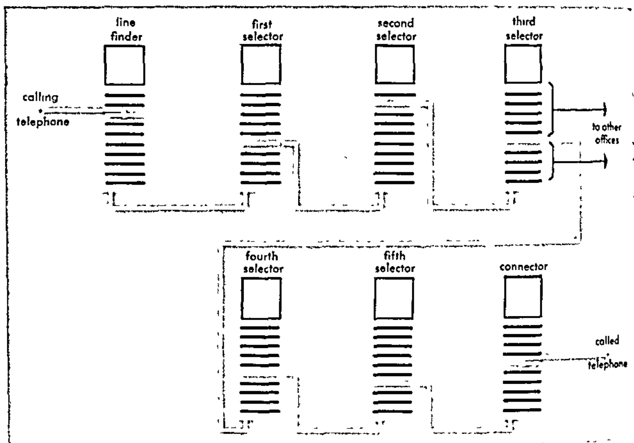


Fig. 10. Diagram of 7-digit step-by-step switching system.

phone number is pulsed over the trunk from the local office to the tandem point.

The crossbar tandem is arranged for 2-wire transmission, that is, the voice is carried in both directions over a single pair of wires

Step-by-step switching system. A system employing a step-by-step switch is described below.

The step-by-step switch, as used in the United States and Canada, was invented by Almon B Strowger. It is a two-motion switch (Fig 9), consisting of a shaft which can be driven step-by-step in a vertical direction and subsequently can be moved step-by-step in a rotary direction. From one to ten steps can be taken in either direction. Attached to the shaft are a set of brushes, or wipers, which make contact with the associated semicircular bank terminals Vertical and rotary magnets are provided to step the shaft by means of a pawl mechanism A release magnet is also provided to cause the switch to return to the normal position at the conclusion of the call Several relays to control its operation are also part of the switch.

Each bank consists of 100 sets of terminals arranged in 10 rows or levels with 10 sets of terminals in each row. The terminals are in the arc of a circle and when the bank is mounted on its switch, the center of the circle is coincident with the center line of the shaft. By lengthening the shaft and providing duplicate brushes and banks, the switch can accommodate 200 sets of terminals.

The switches are of three main types, line finders, selectors, and connectors. The line finder is arranged to step vertically under its own control to the proper level and then step in the rotary direction until it reaches the line that is requesting service. The selector is arranged to step vertically in response to pulses dialed into it and then rotate into the bank until it reaches an idle trunk on that level. The connector is arranged to step in both the vertical and rotary directions in response to pulses dialed into it

Since most central offices are arranged for a maximum of 10,000 main telephones, 4-digit numbers can be used to identify them. A switching system for this number of telephones consists of line finders, two selector stages, and connectors. The first two digits of a number actuate the first and second selectors, and the last two digits operate a connector This type of switching system is adequate for single-office towns These are often known as community dial offices (CDO).

If the office is one of a small group, another selector stage is added and 5-digit numbers are employed in the switching system. As many as eight offices can be in such a network, the digits 1 and 0 not being available for office selection since they are normally used for other purposes.

Many step-by-step central offices are in still larger areas or are arranged for nationwide dialing Hence, 7-digit numbers are used and five selector stages are employed Figure 10 shows such a switching system. All available levels in such a system may not always be needed To avoid the expense of completely equipping unneeded selector stages, these selectors may be arranged for digit absorbing, that is, for abs[illegible] the digit p[illegible]ed without any stepping o[illegible] By [illegible]

he 7-digit numbering plan may be preserved without providing selectors for the codes not in use

The arrangement just described is directly controlled by the pulses dialed by the subscriber. To obtain flexibility in operation, a director has been used in England and (to a lesser degree) in the United States. The director contains equipment for registering the digits dialed by the customer and for pulsing the same or other (substituted) digits to operate the switches ahead. The numbering plan is thereby divorced from the routing plan. A routing through a tandem office, requiring the pulsing of additional digits, may be used to secure economical trunking. Conversely, where switches are not needed for routing purposes, certain pulses may be deleted and the cost of the switches saved. The director may be used for alternate routing and to provide a number of other features, such as special routings for certain classes of subscribers.

The step-by-step system lends itself to the provision of automatic ticketing equipment, permitting subscribers to dial their short and long haul toll calls without the assistance of operators. Data for records and for computing the charges are recorded automatically by the equipment.

Other switching systems. Both the step-by-step system, which is the most common system in use throughout the world, and the crossbar system are made by a number of different manufacturers. Also a number of other systems are in wide use, basically the same in purpose, but using different types of equipment and circuit arrangements. A single-motion rotary switch or uniselector is sometimes used in place of the two-motion switch. One manufacturer of crossbar systems for small offices uses an arrangement in which the crossbar switches are operated by direct control, that is one after another as in the step-by-step system.

Panel system This system has been used mainly in the largest cities of the United States. It is a common control system in which the selectors are power-driven upward to select a trunk or line and downward to return the selector to normal at the conclusion of the call. The terminals of the trunks and lines are located in vertical panels from which the system gets its name Each panel on the intermediate and final selector frames has 100 sets of terminals, and there are five panels on each frame. As a selector is driven upward, pulses are sent back from the selector to the sender in the office in which the call originates This sender counts the pulses sent back and when enough have been counted to reach the desired terminal the upward motion of the selector is stopped Each line finder frame has 10 panels of 40 terminals, accommodating 400 lines per frame

Rotary system Also a power driven common control system, this is used mainly in Europe and in South and Central America The switches are of the unidirectional rotary type. Line finders may be of either 100- or 200-point capacity. Group selectors are usually 300-point, while final selectors are 200-point, although in one of the systems 100-point rotary switches are used throughout. Pulses are used to control the switching functions as in the panel system.

XY system. Made by a number of manufacturers in Europe and the United States, the XY system employs a 100-point, two-motion switch operated magnetically. The switch is flat, and the first or X motion is in a horizontal plane and to the right. The second or Y motion is also horizontal, but at right angles to the first, and carries the brushes into the bank assembly. The system is directly controlled by pulses from the subscriber's dial; the switches are operated one after another as in the step-by-step manner. The flat construction of the switch permits the switches to be stacked one on top of another on the frames, thereby facilitating multiple wiring of the banks

Ericsson 500-line system. This is a power-driven system, utilizing a flat-type switch which is driven first in a rotary direction to one of 25 positions and then in a radial direction to one of 20 sets of terminals. It therefore has access to 500 sets of terminals. See COMMUNICATIONS, ELECTRICAL; TELEPHONY. [C.A.A.]

Switching theory

The theory of circuits made up of ideal digital devices. Included are the theory of circuits and networks for telephone switching, digital computing, digital control, and data processing

Switching theory generally is concerned with circuits made of devices or elements that can be in two or more discrete conditions or states. Examples of such devices are switches or relay contacts, which can be opened or closed, rectifying diodes, which can be either forward- or back-biased, switching tubes or transistors, which can be saturated or cut off, and magnetic cores, which can be magnetized to saturation in either of two directions. Switching theory establishes an ideal representation of the digital circuit, examines the properties of the representation, then interprets these as properties of the circuit Switching theory is not concerned with the physical phenomena of action or stability in a particular condition or with the details of transition from one state to another. It takes these as established and proceeds to examine more or less complex combinations of digital devices whose properties are assumed to be ideal.

The bulk of switching theory is concerned with circuits made of binary (two-valued) devices, since these are most common Switching theory can be based in part on mathematical logic. See BOOLEAN ALGEBRA for a convenient set of concepts and manipulations for the study of certain switching problems

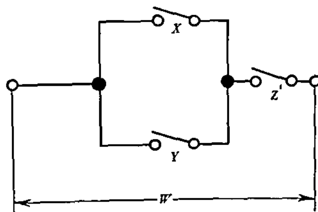
A switch[illegible]
only by [illegible]
tional c[illegible]
cuit in [illegible]
by inpu[illegible] time is called a sequential circuit.

Combinational circuits. A rule by which the outputs of a combinational circuit can be determined from its inputs is called a switching function. Since

the variables are discrete, a switching function may be expressed in tabular form as a truth table, or may be indicated by a diagram or geometric pattern. If the function and variables are binary, the symbols 1 and 0 are commonly used to represent the two values. The function may then be represented by a Boolean algebraic expression. The two values of a switching function can represent closed and open circuits, as for switches or relay contacts, or high and low or plus and minus voltages, as in electronic circuits.

The simplest combinational switching functions are the NOT function, the AND function, and the OR function. The NOT function is designated by the prime in Boolean algebra; $Y = X'$ means that Y is closed (high, plus) when X is open (low, minus), and vice versa. The AND function is designated by the Boolean product; $Z = X \cdot Y$ means that Z is closed (high, plus) only if both X and Y are closed The OR function is designated by the Boolean sum; $Z = X + Y$ means that Z is closed if either X or Y or both are closed. All other combinational switching functions can be made by combining these elementary building blocks.

For example, Fig. 1 shows a switching circuit with three switches, or contacts, X, Y, and Z', each of which can be either open or closed These can be thought of as input variables The circuit as a whole will be open or closed depending upon the individual positions of X, Y, and Z'. Its condition can be designated by W, an output variable. Let 0 represent the open condition, and 1 the closed



X	Y	Z'	W
0	0	1	0
0	0	0	0
0	1	1	1
0	1	0	0
1	0	1	1
1	0	0	0
1	1	1	1
1	1	0	0

Fig 1. Combinational circuit, $W = Z'(X + Y)$. X and Y are normally open contacts. Z' is a normally closed contact.

condition. The table in Fig. 1 represents the switching function of the circuit. The Boolean expression for this function is $W = Z'(X + Y)$. To interpret this expression the rules of simple Boolean algebra must be used.

$$\begin{array}{lll} 0+0=0 & 0\cdot 0=0 & 0'=1 \\ 0+1=1 & 0\cdot 1=0 & 1'=0 \\ 1+0=1 & 1\cdot 0=0 & \\ 1+1=1 & 1\cdot 1=1 & \end{array}$$

Switching theory establishes a number of methods for analysis and synthesis of combinational circuits. A significant problem is minimization, that is, given a switching function, to synthesize the simplest circuit which will realize it. A problem of some theoretical difficulty is that of realizability, that is, given a statement of specifications, to determine whether a switching circuit exists which satisfies them.

Analysis of a series-parallel combination of switches or relay contacts can be carried out by a direct application of Boolean algebra Variables or terms corresponding to contacts, or combinations in parallel, are added, and those in series are multiplied. The values are interpreted according to the rules of Boolean algebra. Similar methods can be applied to combinational circuits which employ diode rectifiers, vacuum tubes, or transistors Circuits that are not series-parallel can be dealt with by an extension of the Boolean method, by the use of matrices with discrete-valued elements, or by a number of special methods

A switching function can be simply synthesized as a series-parallel combination of contacts by giving Boolean symbols circuit interpretations explained previously Electronic logic circuits can be synthesized in a similar fashion. This approach will lead to a method for embodying any switching function expressed in Boolean terms. The Boolean expression of a function given in tabular or diagrammatic form is easily obtained.

Synthesizing the minimal circuit, or minimization, is more difficult since for every switching function there are many possible circuits. Where the number of variables is small, the minimization problem can often be reduced to one that has already been solved Tables of minimal or nearly minimal solutions for relay circuits and vacuum tube circuits are available for circuits with one output and as many as four inputs Harvard chart methods and Karnaugh map methods utilize geometrical relationships to explore systematically functions with one output and as many as six inputs

As the number of variables increases, the possible number of functions rapidly becomes large For example, there are more than 10^{19} different functions of six binary variables. No completely general and practical design methods have been found However, a growing array of special methods for synthesis and minimization is available.

Sequential circuits. Since the outputs of sequential circuits depend on past, as well as present, i

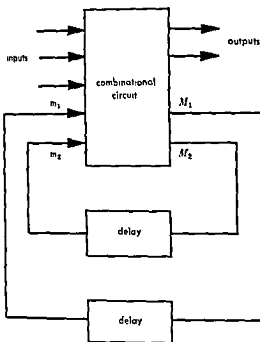


Fig 2 Sequential circuit with two memory loops.

puts, they must contain means for remembering or storing the effect of past inputs, such as locking relays, flip-flops, delay lines, or magnetic cores. A device with two stable states can remember one binary digit, *or bit* The amount of memory in a circuit can be measured either in bits or in internal states. An internal state of a circuit is a particular *configuration* of its internal memory devices. The number *of internal* states is equal to 2^n, where n is its number of bits Binary counters and shift registers are examples of sequential circuits

It is possible to represent a sequential circuit as a combinational circuit *with feedback* Thus, the combinational circuit of Fig 2 becomes a sequential circuit with two bits of memory if two of its outputs are connected *to two of its* inputs Any such closed loop must contain gain and *some* delay, sometimes additional delay is inserted

If the combinational circuit and the delays in Fig 2 are completely specified the *internal* description of the circuit is known and its behavior can be *analyzed. If* the switching function of the combinational circuit is such that $m_1 = M_1$ and $m_2 = M_2$ for a given set of inputs, *no change can* occur as a result of the action of the memory loops and the circuit *is* stable, otherwise, it is unstable. If it is unstable, the inputs must cause a transition to a new state, which in turn may be stable or unstable If no stable *state* is reached, the circuit is said to buzz. If the state to which a circuit may pass *depends* on which of two or more memory loops acts first, *the* circuit is said to have a race condition, and its performance may be ambiguous. This difficulty does not occur in circuits in which changes are caused or timed by repetitive clock pulses. Such circuits are called synchronous. Circuits which make transitions at the natural internal rate are known as asynchronous, and must be designed with greater care.

To proceed from external circuit requirements to an internal description of a sequential circuit may require art and skill, as well as knowledge of switching theory. *See* DATA PROCESSING SYSTEMS; DIGITAL COMPUTER; SAMPLED-DATA CONTROL SYSTEM; SWITCHING CIRCUIT; SWITCHING SYSTEMS (COMMUNICATIONS). [W D L.]

Bibliography: H. H Aiken et al, *Synthesis of Electronic Computing and Control Circuits*, 1951; S. H Caldwell, *Switching Circuits and Logical Design*, 1958; J. T. Culbertson, *Mathematics and Logic for Digital Devices*, 1958; R. A. Higonnet and R. A. Grea, *The Logical Design of Electrical Circuits*, 1958; W. Keister, A. E. Ritchie, and S. H. Washburn, *The Design of Switching Circuits*, 1951.

Switching tubes

There are a large number of tubes which may be used for various switching purposes. Usually ordinary tubes can be used to perform this function, therefore only a relatively small number of special tubes are designed for this purpose. Ordinary vacuum tubes can be used to switch small amounts of current by simply applying a rapid change of voltage to one of the control electrodes. For switching of larger currents, transmitting type vacuum tubes capable of greater power dissipation are needed Switching with vacuum tubes can be *done at rapid* rate, with the changes being effected in microseconds (µsec) or even hundredths of a *microsecond*

For switching of currents larger than several amperes (amp), it is generally necessary to make use of gas tubes, *such* as the thyratron and the ignitron (*see* IGNITRON; THYRATRON).

Signal switching circuits are normally designed to make use of devices that can offer, on demand, either of two *discrete values* of ac resistance to the flow of signal current. Because of simplicity *and* economy, the diode is widely used. It may take *the form of a semicond*

maintained positive relative to the cathode by means of a dc bias *voltage*. Upon reversal of this anode bias voltage, the ac resistance rises toward *infinity See* DIODE, VACUUM.

Where circuit problems and signal levels require isolation and amplification, the triode or multielement vacuum tube may be employed in the same manner as the *diode*. *See* TRIODE, VACUUM; VACUUM TUBE

Of the specialized switching tubes several interesting types are on the market. One, the electron beam

anode receiving the beam *acts* as a closed switch while the other acts as an open switch. Switching *times* of less than 1 µsec are possible

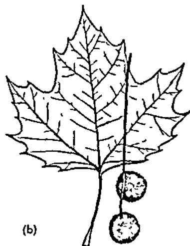
Another high-vacuum switching tube is the magnetron beam-switching tube, in which a series of anodes are equally spaced about a hot cathode. Auxiliary electrodes permit forming an electron beam from the cathode to one of the anodes. A rapid change in potential of one of the auxiliary electrodes will cause the beam to step or transfer to the next electrode, and so forth A magnetic field coaxial with the cathode is employed to cause the beam to transfer. Tubes of this type are commonly fitted with 10 anodes and function as high-speed, decade-counter tubes

Another type of counting-switching tube consists of a gas-filled diode containing 10 anodes arranged in a circle about a cold cathode. Auxiliary electrodes are provided between the anodes to facilitate the transfer of a conducting glow discharge from one anode to another. A change in potential on one of the sets of electrodes or on the cathode will cause the glow discharge to move from one anode to the next, and so forth. Upon the tenth pulse the count would be back to the original anode. *See* COUNTER, DIGITAL; COUNTING CIRCUIT; SWITCHING CIRCUIT [K.R S.]

Sycamore

American sycamore, *Platanus occidentalis*, a member of the plane tree family, known also as American plane tree, buttonball, or buttonwood, and ranging from southern Maine to Nebraska and south into Texas and northern Florida. Ordinarily this tree is 60–120 ft in height and has a trunk which is 2–5 ft in diameter. Individuals 140 ft tall and 14 ft in diameter have been recorded It has the most massive trunk of any American hardwood. Characteristic are the white patches which are exposed when outer layers of the bark slough off, the simple, large, lobed leaves, whose stalks completely cover the conical winter buds; and the spherical fruit heads that are always borne singly in the American species and persist throughout the winter. The tough, coarse-grained wood is difficult to work, but is useful for butchers' blocks, saddle trees, vehicles, tobacco and cigar boxes, crates, and slack cooperage. The trees are usually scattered through the forest in moist soil A rough estimate of the total stand is 3,000 000,000 board feet (bd ft). The average annual cut is less than 50,000,000 bd ft





Sycamores. (a) American sycamore, *Platanus occidentalis*. (b) London plane tree, *Platanus acerifolia* (From A. H. Graves, *Illustrated Guide to Trees and Shrubs*, rev. ed., Harper, 1956)

London plane, *P. acerifolia*, is supposedly a hybrid of *P. occidentalis* and *P. orientalis*, and it is one of the most desirable trees for planting in crowded cities because of its resistance to injury from gases, smoke, dust, and drought. It can be recognized by the usually three-lobed leaves resembling those of a maple, and by the fruit balls borne in groups of two or three. The sycamore of Europe is usually understood to be a maple, *Acer pseudoplatanus*, known also as the sycamore maple. *See* FOREST AND FORESTRY; TREE. [A H G]

Sycettida

An order of calcareous sponges of the subclass Calcaronea This order includes the families Sycettidae, Heteropiidae, Grantiidae, Amphoriscidae, and Lelapiidae Choanocytes occur in flagellated chambers and the spongocoel is not lined with these cells A distinct dermal membrane is present in all but the Sycettidae. The canal system is syconoid to leuconoid. Common genera are *Sycetta*, *Heteropia*, *Amphoriscus*, *Grantia*, and *Leucilla* *See* CALCAREA [C B.C.]

Syenite

A phaneritic (visibly crystalline) plutonic rock with granular texture composed largely of alkali feldspar (orthoclase, microcline, usually perthitic) with subordinate plagioclase (oligoclase) and dark-colored (mafic) minerals (biotite, amphibole and pyroxene). If sodic plagioclase (oligoclase or andesine) exceeds the quantity of alkali feldspar the rock is called monzonite. Monzonites are generally light to medium gray, but syenites are found in a wide variety of colors (gray, green, pink, red), some of which make the material ideal for use as ornamental stone.

Composition. Syenites may be classed as normal (calc-alkali) syenites or alkali syenites. In the latter the alkali feldspar and mafics are soda rich. Intergrowths of potash and soda feldspar (perthite) are of various types and are strikingly developed Feldspar grains usually show fair crystal outlines

(subhedral) or may be irregular (anhedral). Many are highly interlocking Sanidine occurs in some of the finer grained varieties and in syenite porphyry. Normal syenites generally carry crystals of soda plagioclase (oligoclase) which are usually subhedral and may be zoned (with calcic cores and sodic rims). Some alkali syenites contain discrete grains of albite Plagioclase of monzonites may be as calcic as sodic andesine

Black flakes (microscopically brown) of biotite mica and irregular to stubby prisms of green hornblende are characteristic of normal syenite. Diopsidic augite is the most common pyroxene and frequently forms cores within hornblende crystals. In alkali syenite the mafic minerals show wide variation Biotite is deeply colored and iron-rich Amphiboles are soda-rich (arfvedsonite, hastingsite, or riebeckite) and are commonly zoned Diopsidic and titanium-rich augite crystals are commonly encased by shells of aegerine-augite and aegerite.

Minor constituents may include quartz which is usually interstitial. When present in amounts between 5 and 10%, the rock is called quartz syenite, in excess of this amount, the rock becomes a granite Small quantities of feldspathoid (nepheline, sodalite, or leucite) may be present, but if in excess of 10%, the rock becomes a feldspathoidal syenite (nephiline syenite).

Accessory minerals include zircon, sphene apatite, magnetite, and ilmenite. Accessories in special varieties of syenite include iron-rich olivine, corundum, fluorite spinel, and garnet

Texture. The texture of syenite is most commonly even grained Very coarse or pegmatitic textures are local In some syenites numerous, relatively large crystals (phenocrysts) of alkali feldspar give the rock a porphyritic texture These may be of early or late generation and may range from euhedral (well-formed crystals) to anhedral. They are particularly abundant in the finer-grained varieties and in syenite prophyries

Structure. A variety of directive structures may be present Banding and parallel, wavy streaks (schlieren) of different minerals are seen in some syenites, flow structures due to clustering and par- allel orient [illegible]

[illegible] these directive features represent effects of magma flow, in others they represent vestigial bedding or foliation in metasomatic or metamorphic rocks.

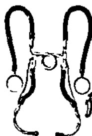
Occurrence and origin. Syenite is an uncommon plutonic rock and usually occurs in relatively small bodies (dikes, sills, stocks, and small irregular plutons). Normal syenite may be associated with monzonite, quartz syenite, and granite, whereas alkali syenites are associated with alkali granites or feldspathoidal rocks

Many syenites have crystallized directly from syenitic magma (rock melt), others may have formed by reaction between magma of nonsyenitic composition and abundant contaminating rock fragments. Still others may have formed metasomatically as alkali-rich emanations, perhaps escaping from deeply buried magmas, permeated rocks of special composition, and replaced them with abundant alkali feldspar. See IGNEOUS ROCKS; MAGMA; METAMORPHISM; METASOMATISM.

[C.A.CA]

Symballophone

A double stethoscope for the comparison and lateralization of sounds, invented in 1937 by W. J. Kerr and colleagues. This device makes use of the functions of the two ears in stereophonic hearing. By means of two chest pieces and tubing leading sound waves from each chest piece to both ears, with tubing 15 cm longer to the opposite ear, a lateralizing effect is achieved through illusion. The human ear can perceive differences of only 0.000032 sec, or the time required for sound to travel only 1 cm



The symballophone. (From O. Glasser, ed., *Medical Physics, Year Book Publ, 1944*)

A symballophone permits the use of the remarkably acute functions of the two ears to compare the intensity and varying quality of sounds arising in the body or mechanical devices, and a more lasting mental registration can be perceived

Physicians have used this device profitably in connection with normal and abnormal sounds arising in the body Location of the site and extent of pneumonic areas in the lungs, murmurs and abnormal sounds in the heart and blood vessels, and roughening in the joints are readily located Murmurs in the blood vessels may be easily timed and their points of origin, direction and other qualities noted. See BIOPHYSICS.

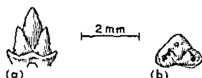
[W.J.KE]

Bibliography O Glasser (ed), *Medical Physics*, 1944

Symmetrodonta

The symm[illegible]

[illegible] to be aberrant triconodonts, they are now recognized as a discrete order of small, carnivorous or insectivorous, primitive mammals The symmetrodont molar consisted of three main cusps which in plan formed a triangle, with the largest cusp forming the apex and the two smaller cusps delimiting the base (see illustration). The teeth were oriented so that the largest cusp of



Lower molar of Tinodon, a symmetrodont. (a) Internal view. (b) Occlusal view. (After G. G. Simpson, 1929)

the upper molars was internal while the largest cusp of the lower molars was external. The long slender jaw did not have a distinct angular process. *See* TRICONODONTA.

The molars of the Jurassic amphiodontid symmetrodonts are distinctly longer than wide and the terminal cusps are so small that functionally the tooth is monocuspid. The molars of the Late Jurassic to Middle Cretaceous spalacotheriid symmetrodonts are shorter, relative to the width, than those of the amphidontids, and the terminal cusps are much larger so that the tooth is functionally tricuspid.

The oldest known symmetrodont probably had a compound lower jaw similar to the jaw of *Morganucodon* (a docodont). Apparently the structure of the jaw was modified during the Jurassic, for the Late Jurassic and Cretaceous symmetrodonts did not have a double jaw articulation. The early symmetrodonts may have been the ancestors of the pantotheres. *See* DOCODONTA; PANTOTHERIA.

[W.A.CL.]

Symmetry laws (physics)

The physical laws which are the expressions of the symmetries existing in the world; a conservation law results from each such symmetry. That is, from each symmetry the existence of a quantity which is conserved (a constant of the motion) can be deduced. Selection rules result from conservation laws. *See* SELECTION RULES (PHYSICS); *see also* NUCLEAR REACTION.

Space-time symmetries. A symmetry (or invariance) of the world exists whenever the description of the laws of physics is unaffected by a change in the frame of reference (*see* FRAME OF REFERENCE). For instance, the position of the origin of a space coordinate system is quite arbitrary, changing it makes no difference in the description of the motion of bodies because the forces between bodies depend only on their relative positions and not on any absolute position. Equivalently, this symmetry expresses itself in that a system of bodies behaves the same if translated to another place. This symmetry of space to translation implies the conservation of momentum. *See* LORENTZ TRANSFORMATIONS; QUANTUM THEORY, NONRELATIVISTIC; QUANTUM THEORY, RELATIVISTIC; RELATIVITY; SPACE-TIME.

Other symmetries of space-time are evidenced by the irrelevance of (1) the origin of the time coordinate, (2) the orientation of a coordinate system in space, and (3) the velocity of a coordinate system (Lorentz invariance). Each of these implies a conservation law, as shown in the accompanying table. All these symmetries are termed continuous because the changes can be arbitrarily small, that is, a finite change can be made bit by bit. The resulting constants of the motion are classical quantities and are additive.

Discrete symmetries (reflections) also exist, for which the irrelevant change is not arbitrarily small. They imply constants of the motion (parities) in quantum mechanics. These parities are multiplicative. *See* PARITY (QUANTUM MECHANICS). For instance, the direction of increasing time is irrelevant; the world is invariant to time reversal (microscopic reversibility). Although macroscopically, future and past seem distinct, this is merely a result of the disposition of matter (a state of anomalously small entropy at some time in the past) in the same way that a point of space seems distinct by having a particular piece of matter there. Space is also symmetrical to reflection of space or to inversion, the reflection of all three directions of space; it is irrelevant which is the positive direction of a space axis or of all three axes. This amounts to the irrelevance of whether a right-handed or a left handed coordinate system is used. The resulting

Invariances (symmetries) and conservation laws

Invariance to	Conserved quantity	Range of validity
Translation of space (homogeneity of space)	Momentum, **p**	Exact
Translation of time (homogeneity of time)	Energy, E	
Rotation of space (isotropy of space)	Angular momentum, **J**	
Lorentz transformation (isotropy of space-time)	Velocity of the center of energy (center of mass)	
Gage transformation	Charge, Q	
Baryon transformation	Net number of baryons	
Lepton transformation	Net number of leptons	
Interchange of identical particles	Symmetry of the wave function (statistics)	
Inversion of time, space, and charge (complete inversion)	CPT	
Reversal of time	Time parity, T	
Reflection of space and charge	Product of parity and charge parity, CP	
Reflection (or inversion) of space	Parity, P	
Reflection of charge (charge conjugation)	Charge parity, C	Violated by the weak interactions
	Third component of isotopic spin, I_3 strangeness, S	
Charge independence of strong interactions (isotropy of isotopic spin space)	Isotopic spin, **I** isotopic parity, G	Violated by the electromagnetic and weak interactions

conserved quantity (eigenvalue of space inversion) is (space) parity. Actually, the preceding statements about space inversion symmetry must be qualified. Although inversion symmetry is observed by the strong interactions (such as nuclear forces) and electromagnetic interactions, it is not observed by the weak interactions such as decays of the elementary particles, including β-decay. The description of a β-decay event depends on the handedness of the coordinate system

Other symmetries. Besides space-time, there are further symmetries in the world. The zero of both scalar and vector electromagnetic potentials is irrelevant; the addition of a constant to an electromagnetic potential is of no consequence (so-called gauge invariance of the first kind) This symmetry implies the conservation of charge. A discrete symmetry can be described as follows: it is *(nearly)* irrelevant which sign of charge is called positive and which is called negative. The qualification "nearly" is necessary here, just as in space inversion, because the weak interactions do not observe the symmetry. Thus the world is *(nearly)* invariant to charge reversal, the interchange of positive and negative charge At first sight this symmetry appears not to exist because one can distinguish positive charge as that which is carried by the heavy constituent of matter, the proton, whereas negative charge is that carried by the light constituent, the electron However, antiprotons (negatively charged protons) and antielectrons (positrons) exist, and a world with (nearly) the same properties as ours would result if all electrons were replaced with positrons and all protons by antiprotons, and in general, if all particles were replaced by their antiparticles (the operation of charge conjugation, *see* ANTIMATTER). The resulting (nearly) conserved quantity is termed charge conjugation parity, or charge parity C More precisely, charge parity is the eigenvalue of the operator of charge reversal (*see* EIGENVALUE) A system can be an eigenstate of charge reversal only if it goes into itself under the operation in other words, is the self-charge conjugate. Such a system must be completely neutral, having no electric or magnetic *moments*; in fact it must have no internal quantum numbers of any kind that change sign under charge conjugation Among the elementary particles this is true only of the neutral π-meson, the photon, and the graviton; their charge parities are the charge parities of their sources, +1, −1, and +1, respectively A self-charge conjugate system of some interest is positronium, a bound state of an electron and a positron, it has $C = (-)^{l+s}$ in a state of orbital angular momentum l and spin s ($s = 0$ or 1) *See* POSITRONIUM, *see also* ELEMENTARY PARTICLE.

CPT Theorem. The reflection symmetries are correlated by the so-called *CPT* Theorem of G. Lüders. This theorem states that a Lorentz invariant field theory is necessarily invariant to the product of the three reflections: charge conjugation C, space inversion P, and time reversal T. Experimentally, it appears that the world (even including the weak interactions) is invariant to time reversal T, and also to the product: charge conjugation times space inversion (CP, or combined inversion).

Invariance of nuclear forces. The nuclear force between two protons is known to be identical to the force between two neutrons (charge symmetry). This means that nuclear forces are invariant to the interchange of protons and neutrons. More generally, the nuclear-force mesons (π-mesons) are to be interchanged also; this charge symmetry operation is $p \leftrightarrow n$, $\pi^+ \leftrightarrow \pi^-$, $\pi^0 \leftrightarrow \pi^0$. The combination of charge conjugation and *the charge* symmetry operation (isotopic inversion, G) carries each π-meson into itself; the π-meson thus has a G parity.

Further, the nuclear force between a neutron and a proton is identical to the force between two protons or two neutrons in the same orbital and spin state (charge independence). *For an extended discussion of nuclear forces, see* NUCLEAR STRUCTURE.

Isotopic spin. The foregoing symmetry can be expressed as the isotropy of a three-dimensional "*isotopic space*" which implies the *conservation of* "angular momentum" or isotopic *spin* I, *in this* space. The component of a particle's *isotopic spin* along the "third" *axis* I_3 is proportional to the *charge of the particle* The consequences of charge *independence are* formally very similar to the *consequences* of the conservation of angular *momentum, except that* nothing here corresponds to *orbital* angular momentum.

In terms of isotopic spin, the charge symmetry operation is a special *rotation* in isotopic spin space, *namely one which* reverses the direction of *the third axis. It follows* from angular momentum *calculus that* a system having $I_3 = 0$ has a charge symmetry parity *which* is $(-)^I$ Thus the charge symmetry *parity of the* π^0 meson is −1. The π-*meson*, *therefore*, has a G parity of −1; nucleonium, a system of nucleon and antinucleon, has $G = (-)^{l+s+i}$ in a state of orbital angular *momentum* l, spin s (0 or 1), and isotopic spin i (0 *or* 1).

All the strong interactions are isotropic in isotopic *space and* conserve isotopic spin; all the strongly *interacting* elementary particles carry isotopic *spin. The* anomalous stability of the heavier (the so-called strange) particles is explained by the fact that charge is conserved, and in both the *strong and* electromagnetic interactions I_3 is conserved (Gell-Mann-Nishijima scheme) *See* ISOTOPIC SPIN; MESON.

Baryonic and leptonic charges. Finally, there are two symmetries which are known only *from* their resulting *conservation laws*, the conservation of baryons *and* the conservation of leptons, also termed the conservation *of* baryonic and leptonic charges These laws *can be* formulated as *the* result of gage invariances, similar to the way in which charge conservation *results* from ordinary gage invariance But there are no *known* field-

in the present cases analogous to the electromagnetic field, and so this "explanation" of the conservations is purely formal and has not led to any fundamental understanding.

Selection rules. As stated earlier, selection rules are an important result of conservation laws; they express whether or not particular reactions can satisfy the conservation laws. A few examples follow.

The conservation of angular momentum, parity, and statistics implies, for example, that unless a level of Be^8 has even angular momentum and positive parity, it cannot decay into two α-particles since they can only be in such states, being identical spinless bosons. *See* QUANTUM STATISTICS.

The conservation of angular momentum and parity implies the selection rules for the emission of radiation. For instance, the selection rules for the emission of electric dipole radiation ($\Delta J = 0, \pm 1$ but not $J = 0$ to another state with $J = 0$) are a consequence of the vector addition rules of angular momentum plus the fact that the electric dipole field is 1^-, that is, its angular momentum is one unit of $\hbar$ and its parity is -1.

The conservation of charge parity C implies that a given state of positronium cannot decay into both an even number and an odd number of photons. Similarly the conservation of isotopic spin parity G implies that a state of nucleonium cannot decay into both an even number and an odd number of π-mesons. Since the weak interactions conserve CP, a neutral K-meson in an eigenstate $CP = -1$ cannot decay into a pair of π-mesons. [C.J.C.]

Bibliography: G. C. Wick, Invariance principles of nuclear physics, *Ann. Rev. Nuclear Sci.*, 8:1–48, 1958; C. N. Yang, Law of parity conservation and other symmetry laws, *Science*, 127(3298):565–569, 1958

Sympathetic nervous system

The portion of the autonomic nervous system, innervating most smooth muscle and glands of the body, which produces a functional state of preparation for flight or combat when stimulated. *See* AUTONOMIC NERVOUS SYSTEM.

Anatomically, this part of the nervous system consists principally of regulatory centers in the thoracic and lumbar regions of the spinal cord. Such centers in the spinal cord contain preganglionic nerve cells which send fibers to the sympathetic trunks or to the great nerve plexuses of the body. The sympathetic trunks, located on either side of the vertebral column, contain postganglionic nerve cells; each trunk is chainlike and lies against the posterior body wall. Additional postganglionic neurons are found in the intertwining networks, or plexuses, of the thoracic, abdominal, and pelvic regions. Peripheral autonomic innervation involves a two-neuron chain as opposed to the direct innervation of voluntary muscle by an axon from a cell in the cerebrospinal axis.

Sympathetic stimulation over this two-neuron system produces alterations in heart rate, bronchiolar size, blood-vessel diameter, glandular secretion,

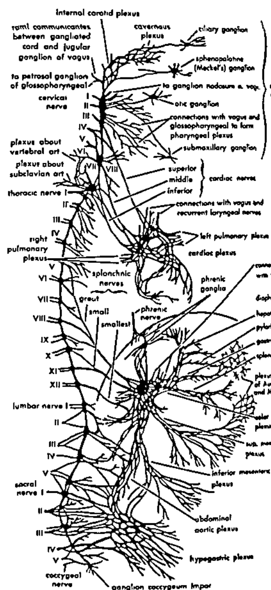
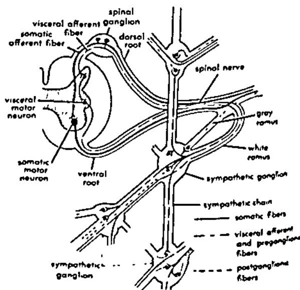


Diagram showing the chain of ganglia and the great plexuses of the sympathetic system. *(After Flower, from J. P. Schaeffer, ed., Morris' Human Anatomy, 11th ed., Blakiston–McGraw-Hill, 1953)*



The visceral reflex arc and sympathetic chain *(From B. A. Houssay et al., Human Physiology, 2d ed., McGraw-Hill, 1955)*

and gastrointestinal activity Most of these structures are also supplied by parasympathetic nerve

OUS SYSTEM.

Some authorities also include sensory mechanisms for the perception of pain, pressure, and the like, in both the sympathetic and parasympathetic systems, but these pathways are largely ill defined

Epinephrine, the active principal of the adrenal medulla, has the same effect on the organism as stimulation of the sympathetic nerves See EPINEPHRINE; NERVOUS SYSTEM. [E.C ST.]

Sympathetic vibration

The driving of a mechanical or acoustical system at its resonant frequency by energy from an adjacent system vibrating at this same frequency. Examples include the vibration of wall panels by sounds issuing from a loudspeaker, vibration of machinery components at specific frequencies as the speed of a motor increases, and the use of tuned air resonators under the bars of a xylophone to enhance the acoustic output Increasing the damping of a vibrating system will decrease the amplitude of its sympathetic vibration but at the same time widen the band of frequencies over which it will partake of sympathetic vibration. See RESONANCE (ACOUSTICS AND MECHANICS); VIBRATION [L E.K.]

Sympetalae

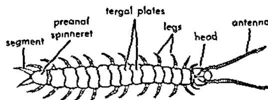
One of the two divisions of the subclass Dicotyledoneae, characterized by flowers with united petals This division contains eight orders or approximately one-fourth of the dicotyledons The other division, archichlamydeae, is made up of 26 orders See ARCHICHLAMYDEAE, see also DICOTYLEDONEAE; EMBRYOPHYTA, PLANT KINGDOM [P D S]

Symphyla

A class of the myriapoda The symphylans like the pauropods, are tiny pale, centipedelike creatures that inhabit humus, soil, or live under debris; in general, they live wherever there is sufficient moisture to preclude excessive water loss. They are similar to the Pauropoda and Diplopoda in being progoneate and anamorphic Their mandibles, like those of ...

...th of the chilopods and lower insects than of the singly maxillate millipedes and pauropods Additional signal characteristics include the following the antennae are unbranched and simple. There is 1 pair of spiracles arising in the head and opening into tracheae, there are 12 pairs of legs, 1 pair per body segment; most of the legs have peculiar basal eversible vesicles with associated styli, the tergites number at least 15 and do not form diplotergites; there is a prominent pair of terminal spinnerets.

As is the case for the Pauropoda, little is known about symphylan biology It is established that they



Symphyla, *Scutigerella immaculata* (Newp), adult Body length up to 7.5 mm. (*From R. E. Snodgrass, A Textbook of Arthropod Anatomy, Cornell University Press, 1952*)

feed upon decaying material as well as upon living plants; their role as greenhouse pests is widely appreciated by agriculturalists Presumably all symphylans hatch with a reduced number of legs, 6–7 in the species investigated, and thereafter undergo molts not only until the adult complement is gained but throughout life, 4–5 years in some forms.

The class consists of three families to which not more than 60 species have been assigned See MYRIAPODA. [R E.C.]

Synanthales

A small order of the plant subclass Monocotyledoneae with one family (Cyclanthaceae) including 6 genera having about 50 species, 35 of which belong to the genus *Carludovica*. These are palmlike, somewhat woody climbers, epiphytes, or shrubs of the American tropics. The leaves of *Carludovica palmata* (jipijapa) are gathered when young and cut into strips to be bleached and woven into Panama hats. See EMBRYOPHYTA; MONOCOTYLEDONEAE; PLANT KINGDOM [P.D S]

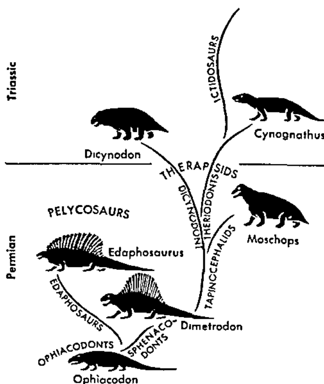
Synapsida

The reptiles grouped under the subclass Synapsida, or mammal-like reptiles The array, in general, is characterized by a temporal fenestra that lies below the junction of the postorbital and squamosal bones, a so-called lower temporal opening. In advanced forms, however, the postorbital-squamosal bridge is absent. Mammals arose from this group of reptiles during the Triassic Period

Synapsids first appear in the geological record in the Upper Carb...

... span of time, they underwent a broad adaptive radiation on land and made minor invasions of aquatic habitats. There were two major phases of this radiation, one early, by pelycosaurs, and the other later (late Permian and Triassic) by therapsids. During each phase both carnivores and herbivores developed Representatives ranged from about one foot to over twenty feet in length.

Irrespective of the nature of their adaptations, evolving stocks became increasingly mammal-like in the course of their history. The most primitive known synapsids were very similar to captorhinomorph reptiles (cotylosaurs), their presumed source. These most primitive forms are known from deltaic deposits of Texas. Here also occur



Evolution of the synapsid or mammal-like reptiles. *(From E. H. Colbert, Evolution of the Vertebrates, Wiley, 1955)*

the several lines of pelycosaurs. Among these, a group called sphenacodonts was progressive and led to the more advanced mammal-like reptiles, the therapsids. Evolution of the therapsids has been traced through rich deposits of their remains in U.S.S.R., South Africa, and South America *See* COTYLOSAURIA

Several developing lines of carnivorous therapsids attained very mammal-like skulls and skeletons, and there is indirect evidence that their soft anatomy and physiology were similarly close to the mammalian level Extremely mammal-like forms compose a group called ictidosaurs. From the highly advanced therapsids and ictidosaurs, mammals came into being in the Late Triassic. Possible ancestral stocks are known from all continents except Australia and Antarctica. The transitions to mammals, which probably were several in number, are so well documented that classification of some genera as reptiles or mammals must be based on arbitrary criteria. *See* ICTIDOSAURIA; PELYCOSAURIA; THERAPSIDA. [E C O]

Synaptic transmission

The mode of transfer of activity at a junction (synapse) between a sensory receptor or a presynaptic neuron and a postsynaptic cell. The latter may be another neuron, or an effector (muscle fiber, electroplaque, or gland cell). Transmission is unidirectional, pre- and postsynaptic functions being distinct. The postsynaptic membrane does not respond to electrical stimuli and is electrically inexcitable. Excitation is presumably by a form of neurosecretion, release of a chemical transmitter substance by the active membrane of presynaptic terminals. Vesicles, thought to contain transmitters and to be released during activity, occur in presynaptic terminals. *See* ELECTRIC ORGAN (BIOLOGY); NERVOUS SYSTEM.

Transmissional activity. The basis of transmissional, as of conductile activity, is a transducer action of the excited membrane, increased permeability to ions, usually leading to change in the membrane polarization (electrogenesis). As electrically inexcitable activity, postsynaptic potentials, excitatory and inhibitory, are distinct from electrically excitable graded responses and spikes. In nerve cells that are diffusely innervated (some muscle fibers, electroplaques, and neurons) transmissional and conductile membranes are intermingled. The former presumably is that opposite presynaptic terminals, but is not distinguishable by present-day anatomical methods At end plates of vertebrate twitch muscle fibers and in some electroplaques, synaptic membrane is limited to one or a few innervated zones. *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY.

Synapses that produce excitatory or inhibitory effects on electrically excitable membrane differ in transducer actions Increased permeability to all ions in activated excitatory membrane leads to decrease or abolition of resting polarization The active state, presumably caused by and lasting during transmitter action, is a period of lowered membrane resistance and inward flow of synaptic current The excitatory postsynaptic potential, which is generated at this time, persists as a depolarization of diminishing amplitude for some time after transmitter action is terminated, the persistence and rate of decay being manifestations of electrotonic spread of potential in the passive network of the membrane. This potential is a stimulus for electrically excitable activity Transducer actions for specific ions, K^+ or Cl^-, or both, characterize three different types of inhibitory synaptic membrane, the cell tending in all three cases to gain negative charge by loss of K^+, entry of Cl^-, or both Inhibitory effects are exerted chiefly during the active phase of the inhibitory postsynaptic potential when the lowered membrane resistance reduces the *IR* drop of the depolarizing potential generated by an excitatory synapse However, depolarization may also be somewhat diminished by the countervailing potential remaining during electrotonic decay of a hyperpolarizing inhibitory potential

Interplays of excitation and inhibition occur in various peripheral neurosensory and neuromuscular complexes, as well as in central nervous systems, even in relatively simple organisms They provide various levels and degrees of error sensing and feedback correction, permitting the organism to exercise a high degree of coordination and precision Activity of cells which lack electrically excitable membrane is initiated by postsynaptic potentials of either variety, which then are neither excitatory nor inhibitory, but only electrical manifestations of activity—discharges of electroplaques, contractions of muscle fibe ecretion in glands.

essage transmission by receptor cells *See* NERVOUS SYSTEM; NERVOUS SYSTEM (INVERTEBRATE).

Drug sensitivity. Pharmacological sensitivity to excitant (synapse activator) or depressant (synapse inactivator) drugs may be specific to depolarizing or hyperpolarizing synapses, but also cuts across that classification, presumably because synaptic membranes of both kinds are activated by the same transmitter agent. Of the synaptically active compounds that occur in animals, only a few are found in and produced by nerve fibers. These alone qualify as possible transmitters The best established are acetylcholine and various catecholamines, which respectively define the cholinergic and adrenergic systems. Axodendritic and axosomatic synapses, initially classified by location on dendrites or cell bodies of neurons, also may be pharmacologically distinguished. Several varieties of invertebrate synapse are insensitive to the common synaptic drugs, indicating as yet unknown transmitter types. *See* EPHAPTIC TRANSMISSION

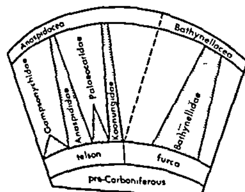
[H GT]

Synbranchiformes

An order of anguilliform fishes that, unlike true eels, have the premaxillae present as distinct bones. This group is also known as the Symbranchii The small gill apertures are confluent across the breast. The gills are poorly developed and respiration is accomplished in part by highly vascularized buccopharyngeal pouches There are no fin spines or pectoral fins and the median fins, if developed, are continuous. Pelvic fins, if present, are small and located on the throat. The body may be naked or scaled There is no swim bladder The group, which has no fossil record, is classified into 2 suborders, 3 families, 7 genera, and 12 species These eel-like fishes inhabit swamps, caves and sluggish fresh and brackish waters of tropical America, Australia, eastern and southeastern Asia, the East Indies, and west Africa *See* ACTINOPTERYGII [R M B]

Syncarida

A superorder of the class Crustacea. There are only a few species of these higher crustaceans They inhabit special regions such as subterranean wells and springs, as well as mountain lakes Their greatest extension was during the Paleozoic era Two orders are recognized, the Anaspidacea and Bathynellacea Phylogenetic relationships within the Syncarida are presented in the illustration. A carapace and oostegites are lacking. Eyes may be pedunculate, sessile, or lacking. All thoracic limbs have exopodites, but none are chelate or subchelate. *See* CRUSTACEA, *see also* ANASPIDACEA; BATHYNELLACEA [H J.]



Phylogenetic relationships within Syncarida

Comparative morphology of Syncarida

Genus	Antennae, number of segments	Thoracic limbs, length of endopodite to exopodite pairs of pleopods telson or furca	Eyes
Gasocaris	1st, numerous, 2nd, numerous with exopodite	Longer, 7, telson	Pedunculate
Campsonychus	1st, numerous 2nd numerous with exopodite	Longer 5; telson	Pedunculate
Anaspides	1st, numerous, 2nd, numerous with exopodite	Longer, 5, telson	Pedunculate
Koonunga	1st numerous 2nd, numerous with exopodite	Longer, 5, telson	Sessile
Bathynella	1st, 7 2nd 7 with exopodite	Longer, 1, furca	Eyeless
Allobathynella	1st, 7 2nd, 5 exopodite lacking	Longer 1, furca	Eyeless
Parabathynella	1st, 7–6 2nd 6–2 present or lacking	Longer, 0, furca	Eyeless
Thermobathynella	1st 6 2nd 5 with exopodite	Longer 0 furca	Eyeless
Brasilibathynella	1st 6, 2nd 6 exopodite lacking	Shorter 0, furca	Eyeless

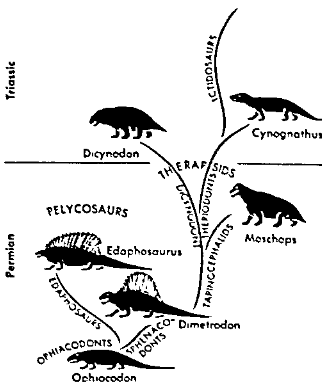
Bibliography P. A Chappuis, *Syncarida* in W Kukenthal and T Krumbach (eds), *Handbuch der Zoologie*, vol 3, pt 1, 1926–1927; H. Jakobi, Biologie, Entwicklungsgeschichte und Systematik von *Bathynella natans* Vejd., *Zool. Jahrb Syst.*, 83 1–63, 1954

Synchro

The name applied to a wide variety of rotary transducers, which are actually synchronous ac motors (or generators) adapted to serve as variable transformers in the measurement of angular position

The construction of a synchro is similar to that of a miniature three-phase synchronous motor or generator The stator, which contains the "three-phase" winding, is a slotted cylindrical structure made up of punched steel laminations, and the rotor contains a single winding, which may be of the salient-pole type, the umbrella type, or the slotted-cylinder type (*see* ALTERNATING-CURRENT GENERATOR)

Instead of exciting the rotor field with a dc current and driving the shaft at a constant velocity as in an alternator, the rotor field is excited with a constant single-phase ac voltage, and the shaft moves at only low speeds, often standing completely still This unit, called a synchro generator,



Evolution of the synapsid or mammal-like reptiles. (*From E. H. Colbert, Evolution of the Vertebrates, Wiley, 1955*)

the several lines of pelycosaurs. Among these, a group called sphenacodonts was progressive and led to the more advanced mammal-like reptiles, the therapsids. Evolution of the therapsids has been traced through rich deposits of their remains in U.S.S.R., South Africa, and South America. *See* COTYLOSAURIA.

Several developing lines of carnivorous therapsids attained very mammal-like skulls and skeletons, and there is indirect evidence that their soft anatomy and physiology were similarly close to the mammalian level. Extremely mammal-like forms compose a group called ictidosaurs. From the highly advanced therapsids and ictidosaurs, mammals came into being in the Late Triassic. Possible ancestral stocks are known from all continents except Australia and Antarctica. The transitions to mammals, which probably were several in number, are so well documented that classification of some genera as reptiles or mammals must be based on arbitrary criteria. *See* ICTIDOSAURIA; PELYCOSAURIA; THERAPSIDA. [E.C.O.]

Synaptic transmission

The mode of transfer of activity at a junction (synapse) between a sensory receptor or a presynaptic neuron and a postsynaptic cell. The latter may be another neuron, or an effector (muscle fiber, electroplaque, or gland cell). Transmission is unidirectional, pre- and postsynaptic functions being distinct. The postsynaptic membrane does not respond to electrical stimuli and is electrically inexcitable. Excitation is presumably by a form of neurosecretion, release of a chemical transmitter substance by the active membrane of presynaptic terminals. Vesicles, thought to contain transmitter and to be released during activity, occur in presynaptic terminals. *See* ELECTRIC ORGAN (BIOLOGY); NERVOUS SYSTEM.

Transmissional activity. The basis of transmissional, as of conductile activity, is a transducer action of the excited membrane, increased permeability to ions, usually leading to change in the membrane polarization (electrogenesis). As electrically inexcitable activity, postsynaptic potentials, excitatory and inhibitory, are distinct from electrically excitable graded responses and spikes. In nerve cells that are diffusely innervated (some muscle fibers, electroplaques, and neurons) transmissional and conductile membranes are intermingled. The former presumably is that opposite presynaptic terminals, but is not distinguishable by present-day anatomical methods. At end plates of vertebrate twitch muscle fibers and in some electroplaques, synaptic membrane is limited to one or a few innervated zones. *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY.

Synapses that produce excitatory or inhibitory effects on electrically excitable membrane differ in transducer actions. Increased permeability to all ions in activated excitatory membrane leads to decrease or abolition of resting polarization. The active state, presumably caused by and lasting during transmitter action, is a period of lowered membrane resistance and inward flow of synaptic current. The excitatory postsynaptic potential, which is generated at this time, persists as a depolarization of diminishing amplitude for some time after transmitter action is terminated, the persistence and rate of decay being manifestations of electrotonic spread of potential in the passive network of the membrane. This potential is a stimulus for electrically excitable activity. Transducer actions for specific ions, K^+ or Cl^-, or both, characterize three different types of inhibitory synaptic membrane, the cell tending in all three cases to gain negative charge by loss of K^+, entry of Cl^-, or both. Inhibitory effects are exerted chiefly during the active phase of the inhibitory postsynaptic potential when the lowered membrane resistance reduces the *IR* drop of the depolarizing potential generated by an excitatory synapse. However, depolarization may also be somewhat diminished by the countervailing potential remaining during electrotonic decay of a hyperpolarizing inhibitory potential.

Interplays of excitation and inhibition occur in various peripheral neurosensory and neuromuscular complexes, as well as in central nervous systems, even in relatively simple organisms. They provide various levels and degrees of error sensing and feedback correction, permitting the organism to exercise a high degree of coordination and precision. Activity of cells which lack electrically excitable membrane is initiated by postsynaptic potentials of either variety, which then are neither excitatory nor inhibitory, but only electrical manifestations of activity—discharges of electroplaques, contractions of muscle fibe secretion in glands

message transmission by receptor cells *See* NERVOUS SYSTEM; NERVOUS SYSTEM (INVERTEBRATE).

Drug sensitivity. Pharmacological sensitivity to excitant (synapse activator) or depressant (synapse inactivator) drugs may be specific to depolar- [illegible] synapses, but also cuts

compounds that occur in animals, only a few are d in and produced by nerve fibers. These alone ify as possible transmitters. The best estab- d are acetylcholine and various catecholamines, h respectively define the cholinergic and ad- rgic systems Axodendritic and axosomatic pses, initially classified by location on den- s or cell bodies of neurons, also may be macologically distinguished Several varieties ivertebrate synapse are insensitive to the com- synaptic drugs, indicating as yet unknown smitter types *See* EPHAPTIC TRANSMISSION

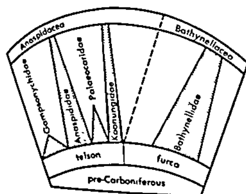
[H GT.]

ıbranchiformes

order of anguilliform fishes that, unlike true , have the premaxillae present as distinct bones. group is also known as the Symbranchii The ll gill apertures are confluent across the breast. gills are poorly developed and respiration is omplished in part by highly vascularized bucco- ryngeal pouches. There are no fin spines or toral fins and the median fins, if developed, are itinuous. Pelvic fins, if present, are small and ated on the throat. The body may be naked or lled There is no swim bladder. The group, which no fossil record, is classified into 2 suborders, amilies, 7 genera, and 12 species These eel-like hes inhabit swamps, caves, and sluggish fresh d brackish waters of tropical America, Australia, stern and southeastern Asia, the East Indies, and st Africa *See* ACTINOPTERYGII [R M B.]

ncarida

superorder of the class Crustacea There are only few species of these higher crustaceans They in- bit special regions such as subterranean wells



Phylogenetic relationships within Syncarida

Comparative morphology of Syncarida

Genus	Antennae, number of segments	Thoracic limbs length of endopodite to exopodite, pairs of pleopods telson or furca	Eyes
Gasocaris	1st, numerous 2nd numerous with exopodite	Longer, 7, telson	Pedunculate
Gampsonychus	1st numerous, 2nd, numerous with exopodite	Longer, 5, telson	Pedunculate
Anaspides	1st, numerous 2nd numerous with exopodite	Longer, 5, telson	Pedunculate
Koonunga	1st, numerous 2nd numerous with exopodite	Longer, 5, telson	Sessile
Bathynella	1st, 7, 2nd, 7 with exopodite	Longer 1 furca	Eyeless
Allobathynella	1st 7 2nd 5 exopodite lacking	Longer, 1, furca	Eyeless
Parabathynella	1st 7–6, 2nd 6–2 present or lacking	Longer 0 furca	Eyeless
Thermobathynella	1st 6 2nd 5 with exopodite	Longer 0 furca	Eyeless
Brasilibathynella	1st 6 2nd, 6 exopodite lacking	Shorter, 0, furca	Eyeless

and springs, as well as mountain lakes Their greatest extension was during the Paleozoic era Two orders are recognized, the Anaspidacea and Bathynellacea. Phylogenetic relationships within the Syncarida are presented in the illustration. A carapace and oostegites are lacking Eyes may be pedunculate, sessile, or lacking All thoracic limbs have exopodites, but none are chelate or subchelate. *See* CRUSTACEA; *see also* ANASPIDACEA; BATHYNELLACEA [H J.]

Bibliography: P A. Chappuis, *Syncarida* in W Kükenthal and T. Krumbach (eds), *Handbuch der Zoologie*, vol. 3, pt. 1, 1926–1927; H. Jakobi, Biologie, Entwicklungsgeschichte und Systematik von *Bathynella natans* Vejd , *Zool. Jahrb Syst.*, 83:1–63, 1954

Synchro

The name applied to a wide variety of rotary transducers, which are actually synchronous ac motors (or generators) adapted to serve as variable transformers in the measurement of angular position

The construction of a synchro is similar to that of a miniature three-phase synchronous motor or generator. The stator, which contains the "three-phase" winding, is a slotted cylindrical structure made up of punched steel laminations, and the rotor contains a single winding, which may be of the salient-pole type, the umbrella type, or the slotted-cylinder type (*see* ALTERNATING-CURRENT GENERATOR)

Instead of exciting the rotor field with a dc current and driving the shaft at a constant velocity as in an alternator, the rotor field is excited with a constant single-phase ac voltage, and the shaft moves at only low speeds, often standing completely still This unit, called a synchro generator,

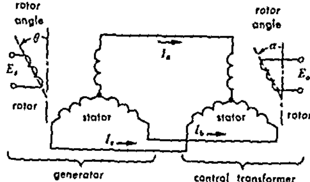


Fig 1. Synchro error-detector system

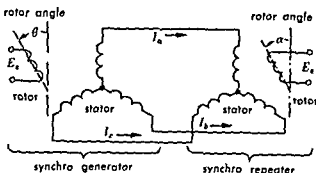


Fig 2. Synchro repeater system

is basically a transformer with one primary (the rotor winding) and three secondaries (the Y-connected windings of the stator). The voltages induced in the secondary windings are proportional to the cosines of the angles between each stator coil and the rotor. Thus an electrical reference frame is formed which may be read at a remote point by means of another synchro, called a synchro control transformer, operating in reverse fashion as shown in Fig. 1

The synchro is completely single-phase, the term three-phase winding is a misleading carryover from the field of ac power machinery to describe a stator that is made up of three fields oriented at 120° to each other

The voltage E_o in Fig. 1 is a maximum when the control transformer rotor angle α coincides with the generator rotor angle θ. To achieve a null condition, the control transformer rotor position must be at an angle of 90° with respect to the generator rotor angle. Thus if the control transformer rotor position is defined as $\beta = \alpha + 90°$, then β will be equal to θ when the output voltage E_o is at a minimum (theoretically zero). Two synchros may be used in this way as an error detector system, with the output voltage E_o representing the error $(\theta - \beta)$.

The best null condition attainable for E_o is limited by quadrature effects resulting from extraneous phase shifts between the various transformer windings due to slight differences in impedance characteristics. Special care must be exercised in the manufacture of synchros to attain good performance When the rotors of synchros are driven at high speeds, extraneous voltages are induced by generator action, and the output signal is not an accurate representation of rotor-angle difference

Two-speed synchro system. This is a method of employing two sets of synchros for error detection in a servomechanism, one set operating through step-up gearing so that they rotate at some integral multiple of the speed of the others. The nonstepped-up synchros are employed for error detection when the error is large, and the stepped-up synchros are switched in to provide more accurate error detection when the error is small. Complications arise from the many possible false null positions when the stepped-up synchros are in operation, and circuits must ensure that the system drives to the correct null position.

Repeater synchro. In a repeater synchro system the rotor winding of the second synchro is excited with the same voltage and frequency as the rotor of the synchro generator, as shown in Fig. 2. When the synchro repeater rotor angle does not coincide with the generator rotor angle, an electromagnetically induced torque drives the repeater rotor toward the position where its angle will correspond with the generator rotor angle A synchro repeater is well damped to avoid excessive oscillation in finding its correct position; otherwise it is identical with a control transformer. This system is sometimes used as a combined error detector and torque motor because the torque induced in the repeater rotor is proportional to the error angle $(\theta - \alpha)$ for small errors These are com

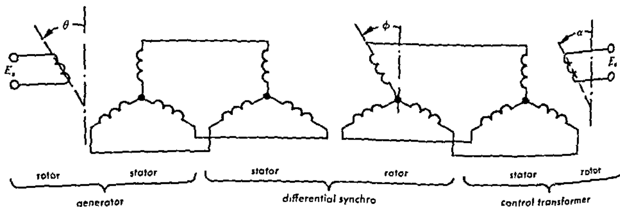


Fig 3. System with differential synchro.

monly used to position remote indicators. *See* REPEATER, SYNCHRO.

Differential synchro. This synchro contains three sets of windings on both the rotor and the stator. When this unit is connected with a synchro generator and control transformer, as shown in Fig 3, the output of the control transformer reaches its null when the angle $\beta = \alpha + 90°$ is equal to the difference between the generator angle θ and the differential synchro rotor angle ϕ. Thus the differential synchro makes it possible to introduce another reference angle into the error detector system

Resolver synchro. This is similar to a synchro generator in construction, but the stator contains ly two windings oriented at 90° relative to each ier, and they are employed to resolve rotor posi-n into sine and cosine component voltage sig-ls Resolver synchros are used in computing vomechanisms and other electromechanical com-ters *See* ANALOG COMPUTER; SERVOMECHANISM. [J L.SH.]

nchrocyclotron

cyclotron for accelerating protons, deuterons, or particles, in which the *frequency of the accelerat-*; voltage is modulated to maintain synchronism th the frequency of the *particle which is spiral-*; out to energies where the relativistic *mass in-*case becomes significant. In Russian literature is device is sometimes termed a *synchrophaso-*n The principle of phase stability controls the dius of the particle to maintain the synchronism the frequency is decreased. For an extended dis-ssion see PARTICLE ACCELERATOR [W K H.P.]

nchronization

ie process of maintaining one operation in step th another. The commonest example is the elec-c clock, whose motor rotates at some integral ultiple or submultiple of the speed of the alter-tor in the power station In television, synchro-zation is essential in order that the electron ams of receiver picture tubes are at exactly the me spot *on the* screen at each instant as is the am in the *television* camera tube at the trans-itter. Synchronism in television is achieved by ansmitting a synchronizing pulse at the end of ch scanning line, to make all receivers move multaneously *to* the start of the next line A simi-r vertical synchronizing pulse is transmitted hen the camera beam reaches the bottom of the cture, to make all beams go back to the top for e start of the next field *See* OSCILLOSCOPE, CATHODE RAY; TELEVISION [J MR.]

ynchronous condenser

synchronous motor that is operated without a echanical l ... er power ... t the lag ... cause it ... give the required amount of correction as induction motors are switched on or off and their loads are changed. Industrial plants use synchronous condensers to improve or completely correct overall power factor, so as to avoid power company rate penalties for drawing highly reactive power Synchronous condensers also help to maintain constant line voltage *See* SYNCHRONOUS MOTOR [J.MR]

Synchronous motor

An alternating-current (ac) motor which operates at a fixed synchronous speed proportional to the frequency of the applied ac power. A synchronous machine may operate as a generator, motor, or condenser depending only on its applied shaft torque (whether positive, negative, or zero) and its excitation There is no fundamental difference in the theory, design, or construction of a machine intended for any of these roles, although certain design features are stressed for each of them. In use, the machine may change its role *from instant to instant.* For these reasons it is preferable not to set up separate theories for synchronous generators, motors, and condensers, but rather to establish a general theory which is applicable to all three and in which the distinction between them is merely a difference in the direction of the currents and the sign of the torque angles.

Basic theory. A single-phase, two-pole synchronous machine is shown in Fig. 1*a*. The coil is on the pole axis at time $t = 0$ and the sinusoidally distributed flux ϕ linked with the coil at any instant is

$$\phi = \Phi_{max} \cos \omega t \quad (1)$$

where ωt is the angular displacement of the coil and Φ_{max} is the maximum value of the flux. This flux will induce in a coil of N turns an instantaneous voltage

$$e = -N\frac{d\phi}{dt} = \omega N\Phi_{max} \sin \omega t = E_{max} \sin \omega t \quad (2)$$

The effective (rms) value of this voltage is

$$E = \frac{E_{max}}{\sqrt{2}} = \sqrt{2}\,\pi f N\Phi_{max} = 4.44 f N\Phi_{max} \quad (3)$$

If the impedance of the coil and its external circuit of resistance R and reactance X is

$$\mathbf{Z} = R \pm jX = Z/\underline{\pm\theta} \quad (4)$$

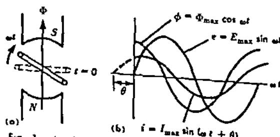


Fig 1. (a, b) Single-phase, 2-pole synchronous machine

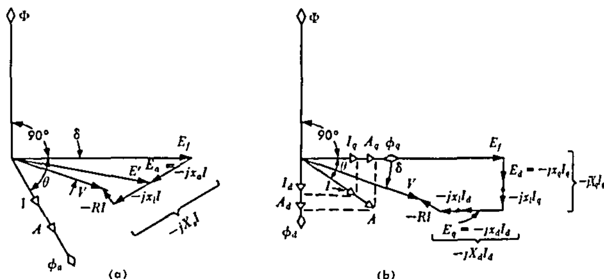


Fig. 2. Vector diagrams of synchronous generators (a) Smooth-rotor machine (b) Salient pole machine.

there will flow a current

$$\mathbf{I} = \frac{\mathbf{E}}{\mathbf{Z}} = \frac{E}{Z}\underline{/\mp\theta} \qquad (5)$$

in which the phase angle θ is taken positive for a leading current. This current will develop a sinusoidal space distribution of armature reaction

$$A = 0.8NI_{max} \sin(\omega t + \theta) \qquad (6)$$

If this single-phase mmf is expressed as a space vector and resolved into direct (in line with the pole axis) A_d and quadrature A_q components it is

$$\begin{aligned} \mathbf{A} &= A_d + jA_q \\ &= 0.4NI_{max}\{[\sin\theta + \sin(2\omega t + \theta)] \\ &\qquad + j[\cos\theta - \cos(2\omega t + \theta)]\} \end{aligned} \qquad (7)$$

In a three-phase machine with balanced currents, the phase currents are

$$\begin{aligned} i_a &= I_{max}\sin(\omega t + \theta) \\ i_b &= I_{max}\sin(\omega t + \theta - 120°) \\ i_c &= I_{max}\sin(\omega t + \theta - 240°) \end{aligned} \qquad (8)$$

Upon writing Eq. (7) for ωt, $\omega t - 120°$, and $\omega t - 240°$ respectively and adding, there results for the polyphase armature reaction

$$\mathbf{A} = A_d + jA_q = 1.2NI_{max}(\sin\theta + j\cos\theta) \qquad (9)$$

The three-phase power of the machine is

$$P = 3EI\cos\theta \qquad (10)$$

and the developed torque is

$$T = \frac{P}{\omega} = \frac{3}{\omega}EI\cos\theta \qquad (11)$$

The above equations constitute the essential description of the synchronous generator. The same equations apply for a motor if the currents are reversed, that is, by changing the sign of the current I. We next interpret these equations in the form of vector diagrams, and recognize the two cases of a smooth-rotor and a salient-pole machine.

Smooth-rotor synchronous machine. In the smooth-rotor machine, the reluctance of the magnetic path is essentially the same in either the direct or quadrature axes. In Fig 2*a* let the flux Φ be selected as reference vector and drawn vertically. Then comparing Eqs. (1) and (2) it is seen that the induced voltage E_f lags the flux by 90°. By Eq. (5) the current I lags the voltage by an angle θ for an inductive circuit, and by Eq (9) causes a constant mmf of armature reaction A in phase with the current. This armature reaction causes a flux ϕ_a, stationary in space with respect to the field poles, which in turn induces a voltage E_a lagging it by 90°. The two induced voltages E_f (due to the field flux Φ) and E_a (due to the armature reaction flux ϕ_a) combine vectorially to give the resultant voltage E'. But the terminal voltage V is less than E' by the resistance and reactance drops, $R\mathbf{I}$ and $jx_l\mathbf{I}$ in the winding, thus

$$\mathbf{V} = \mathbf{E}' - (R + jx_l)\mathbf{I} \qquad (12)$$

The leakage reactance drop $jx_l\mathbf{I}$ lags the current by 90° as does the armature reaction voltage E_a. If a fictitious reactance of armature reaction x_a is introduced to account for E_a it is obvious that Eq (12) may be rewritten as

$$\begin{aligned} \mathbf{V} &= \mathbf{E}_f - jx_a\mathbf{I} - (R + jx_l)\mathbf{I} \\ &= \mathbf{E}_f - R\mathbf{I} - j(x_a + x_l)\mathbf{I} \\ &= \mathbf{E}_f - (R + jX_s)\mathbf{I} \end{aligned} \qquad (13)$$

in which $X_s = x_a + x_l$ is called the synchronous reactance of the machine.

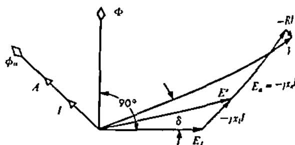
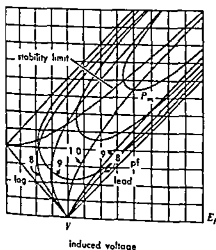


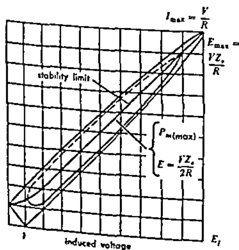
Fig 3. Vector diagram of synchronous motor.

Salient-pole synchronous machine. In a similar fashion the vector diagram for a salient-pole machine, Fig 2b, may be set up, where the effects of saliency result in proportionately different armature reaction fluxes in the direct and quadrature axes, thereby necessitating corresponding direct, X_d, and quadrature, X_q, components of the synchronous reactance

The angle δ in the vector diagrams of Fig. 2 is called the torque angle It is the angle between the field induced voltage E_f and the terminal voltage V and is positive when E_f is ahead of V.

The foregoing equations and vector diagrams were established for a generator. A motor may be regarded as a generator in which the power component of the current is reversed 180°, that is, becomes an input instead of an output current The motor vector diagram is shown in Fig 3 Here the torque angle δ is reversed since V is ahead of E_f in a motor (it was behind in the generator). Therefore a motor d[illegible]





4. V curves (armature current vs induced voltage) synchronous motor.

result the power input, Eq. (10), for a motor is negative, or has become a power output, and the torque is reversed in sign.

When the current I is 90° out of phase with the terminal voltage V the torque angle δ is nearly zero, being just sufficient to account for the power lost in the resistance.

Therefore, a synchronous machine is a generator, motor, or condenser depending on whether its torque angle δ is positive, negative, or zero. For these conditions the output current is respectively at an angle less than ±90°, greater than ±90°, or essentially ±90° with respect to the terminal voltage. Furthermore, depending on this power-factor angle, the field induced voltage E_f may be greater (overexcited) or less (underexcited) than the terminal voltage V, and the machine may be made to take either leading or lagging currents.

Synchronous condenser. A synchronous condenser can be made to draw a leading current and to behave like a capacitance, by overexciting its field Or, it will draw a lagging current on underexcitation. This characteristic thus presents the possibility of power-factor correction of a power system by adjusting the field excitation. A machine so employed at the end of a transmission line permits a wide range of voltage regulation for the line. One used in a factory permits the power factor of the load to be corrected. Of course a synchronous motor can also be used for power-factor correction, but since it must also carry the load current, its power-factor correction capabilities are more limited than for the synchronous condenser.

Power equations. The power P_g and reactance power Q_g of a round-rotor synchronous generator is given by the equation

$$P_g + jQ_g = \left(\frac{VE_f}{Z_s}\sin(\delta - \alpha) + \frac{RE_f^2}{Z_s^2}\right) + j\left(\frac{VE_f}{Z_s}\cos(\delta - \alpha) - \frac{X_sE_f^2}{Z_s^2}\right) \quad (14)$$

in which $\tan \alpha = R/X_s$ and $\mathbf{Z}_s = R + jX_s$

For a round-rotor motor the torque angle δ is negative and the gross mechanical power output (including windage and friction) is

$$P_m = \frac{VE_f}{Z_s}\sin(\delta + \alpha) - \frac{RE_f^2}{Z_s^2} \cong \frac{VE_f}{Z_s}\sin\delta \quad (15)$$

For a salient-pole machine, neglecting resistance, Z_s is equal to the direct-axis reactance X_d, α is zero, and

$$P_m = \frac{VE_f}{X_d}\sin\delta + V^2\frac{X_d - X_q}{2X_dX_q}\sin 2\delta \quad (16)$$

Thus the power or torque depends essentially on the product of the terminal and induced voltages and sine of the torque angle δ; but in the case of the salient-pole machine there is also a second harmonic term which is independent of the excitation voltage E_f This term, the so-called reluctance power, vanishes for nonsaliency when $X_d = X_q$. The small synchronous motors used in some electric

clocks and other low-torque applications depend solely on this reluctance torque. *See* Reluctance motor.

Excitation characteristics. The so-called V curves of a synchronous motor are curves of armature current plotted against field current with power output as parameter. Usually a second set of curves with input power factor (pf) as parameter is superimposed on the same plot. Such curves, Fig. 4, where armature current is plotted against generated voltage, can be determined from design calculations, or from test, and yield a considerable amount of data on the performance of the motor. Thus, given any two of the four variables E_f, I, pf, P, the remaining two may be easily determined, as well as the conditions of maximum power, constant pf, minimum excitation, stability limit, and so forth.

Circle diagrams. The voltage equation (13) and the current equation (5) can be combined in such a fashion as to yield

$$I^2 = \frac{V^2}{Z_s^2} + \frac{E_f^2}{Z_s^2} - 2\frac{V}{Z_s}\frac{E_f}{Z_s}\cos\delta \qquad (17)$$

which is the equation of a set of circles with offset center and with different radii (E_f/Z_s) The locus of these circles is the current

A companion set of circles can be developed giving the locus of I as a function of its power-factor angle for different values of constant developed power.

These two sets of circles are shown in Fig. 5 Such circle diagrams relate the power, pf angle, armature current, torque angle, and excitation.

Fig 5. Circle diagram of synchronous motor.

Losses and efficiency. The losses in a synchronous motor comprise the copper losses in the field, armature, and amortisseur windings; the exciter and rheostat losses of the excitation system; the core loss due to hysteresis and eddy currents in the armature core and teeth and in the pole face; the stray loss due to skin effect in conductors; and the mechanical losses due to windage and friction. The efficiency of the motor then is

$$\text{Eff} = \frac{\text{output}}{\text{input}} = \frac{\text{output}}{\text{output} + \text{losses}} \qquad (18)$$

Mechanical oscillations. A synchronous motor subjected to sudden changes of load, or when driving a load having a variable torque (for example, a reciprocating compressor), may oscillate about its mean synchronous speed. Under these conditions the torque angle δ does not remain fixed, but varies. As a result four separate torques act on the machine rotor:

$$\begin{pmatrix}\text{Synchronous}\\ \text{motor torque}\\ \text{Eq (16)}\end{pmatrix} + \begin{pmatrix}\text{induction motor}\\ \text{torque of}\\ \text{amortisseur}\end{pmatrix} = \begin{pmatrix}\text{torque to}\\ \text{overcome}\\ \text{inertia}\end{pmatrix} + \begin{pmatrix}\text{torque}\\ \text{required}\\ \text{by the load}\end{pmatrix} \qquad (19)$$

The possibility exists that cumulative oscillations *will build up and cause the motor to* fall out of step

Starting of synchronous motors. Synchronous motors are provided with an amortisseur (squirrel cage) winding embedded in the face of the field poles This winding serves the double purpose of motor starting and the limiting of oscillations or hunting. During starting the field winding is either closed through a resistance, short-circuited, or opened at several points to avoid dangerous induced voltages. The amortisseur winding acts exactly as the squirrel-cage winding in an induction motor and accelerates the motor to nearly synchronous speed When near synchronous speed the field is excited and the synchronous torque pulls the motor into synchronism. During starting Eq (19) applies, since all four types of torque may be present. Of course, up to the instant when the field is excited the portion of the synchronous motor torque depending on E_f does not exist, although the reluctance torque will be active.

Other methods of starting have been used If the exciter is direct-connected and a dc source of power is available, it may be used to start the synchronous motor In the so-called supersynchronous motor the stator is able to rotate in bearings of its own, and is provided with a brake band For starting, the stator brake band is released and the stator allowed to come up to nearly synchronous speed by virtue of the amortisseur windings, the field is then excited and the stator brought to synchronous speed, the rotor remaining stationary Then as the brake band is tightened, the torque on the rotor

uses it to accelerate while the speed of the stator correspondingly slackens and finally the stator comes to rest and is locked by the brake band. In this way maximum synchronous motor torque is made available for acceleration of the load. For other types of synchronous motors see HYSTERESIS MOTOR, RELUCTANCE MOTOR. [L.V.B.]

Bibliography: L. V. Bewley, *Alternating-current Machinery*, 1949; A. S. Langsdorf, *Theory of Alternating-current Machinery*, 2d ed., 1955; M. Liwschitz-Garik and C. C. Whipple, *Electric Machinery: A-C Machines*, vol. 2, 1946; A. F. Puchstein, T. C. Lloyd, and A. G. Conrad, *Alternating-current Machines*, 3d ed., 1954.

Synchronous speed

The speed of an alternating-current (ac) machine at which the rotor speed and the frequency of the ac wave are exactly proportional. This is also the speed of the rotating field of an induction machine. In a machine with p poles, the synchronous speed n_s is

$$n_s = \frac{120f}{p} \quad \text{rpm}$$

where f is the frequency of the ac wave. [A.F.P.]

Synchroscope

An instrument used for indicating whether two alternating-current (ac) generators or other ac voltage sources are synchronized in time phase with each other. In one type, for example, the position of a continuously rotatable pointer indicates the instantaneous phase difference between the two sources at each instant, the speed of rotation of the pointer corresponds to the frequency difference between the sources, while the direction of rotation indicates which source is higher in frequency. In more modern synchroscopes, a cathode-ray tube serves as the indicating means.

The term synchroscope is also applied to a special type of cathode-ray oscilloscope designed for observing extremely short pulses, using fast sweeps synchronized with the signal to be observed. See ELECTRIC POWER GENERATION; OSCILLOSCOPE, CATHODE-RAY. [J.MR.]

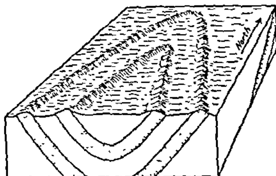
Synchrotron

A device for accelerating electrons or protons in circular orbits in a time-varying magnetic field at nearly constant radius. The magnetic field has a radial gradient to produce focusing. Acceleration is produced by an electrode system driven at a frequency (constant in the case of electrons, variable in the case of protons) chosen so as to keep the orbit radius constant. The orbit radius will adjust to a value maintaining synchronism between the frequency of revolution of the particle in its orbit and the frequency of the accelerating voltage; this action is called phase stability.

An alternating-gradient synchrotron is a synchrotron which employs a radial gradient of the magnetic guide-field alternating in sign as a means of focusing the particle beam. For an extended discussion see PARTICLE ACCELERATOR. [W.K.H.P.]

Syncline

A fold in which the beds are inclined down and toward the axis. Synclines may be symmetrical, asymmetrical, overturned, or recumbent. Most have elongate trends, with axes that plunge from the extremities toward interior points along the axes. Others, called basins, have no distinct trend. In general, the stratigraphically younger beds are found toward the center of curvature, but in complexly deformed regions such simple concepts may not apply.



Block diagram showing relation between structural syncline and topography. The topography indicates that the syncline plunges south. (*From M. P. Billings, Structural Geology, 2d ed., Prentice-Hall, 1954*)

Stratigraphic synclines are those folds, regardless of their observed forms, which are inferred from stratigraphic data to have been synclines originally. Structural synclines are those which have synclinal form regardless of their stratigraphic relations. See ANTICLINE; FOLD AND FOLD SYSTEMS. [P.H.O.]

Syngamy

A process involving the union of sexual cells, or gametes. Customarily these are eggs (female) and spermatozoa (male), and syngamy involves the fusion of a single spermatozoon with a single egg. As such, syngamy is synonymous with the word fertilization. Accompanying the union of sexual cells are rather profound physiological and structural changes within the egg, but the process is considered terminated when a fusion of the gametic nuclei takes place. Syngamy in the ultimate sense of nuclear fusion was first seen in animal cells by O. Hertwig and H. Fol in 1874, and in plant cells by E. Strasburger a year later. See FERTILIZATION.

Syngamy bears a significant relation to *genetics* in that in effect it is the opposite of meiosis. Meiosis reduces, and generally halves, the chromosome number of a diploid cell to produce haploid gametes; *syngamy*, through the fusion of haploid nuclei, restores the diploid chromosome number. Paternal

and maternal genes are thereby brought together in the resulting offspring developing from the fertilized egg or zygote. *See* GENETICS; MEIOSIS.

Syngamy bears a further relation to sexual reproduction in that entry of the sperm into the egg activates the egg, leading to a cycle of cell division and differentiation which eventually leads to the formation of a new individual. The life cycle of any sexual organism is, therefore, an alternation of diploid and haploid phases, with meiosis and synga [illegible]

ga [illegible] d haploid phases can vary widely in relation to the entire life cycle. In many lower forms, however, it is not always possible to speak of syngamy as being the union of a spermatozoon and an egg, since the gametes are similar in appearance and undifferentiated as to sex. *See* CELL DIVISION. [C.P.SW.]

Syphilis

A subacute and chronic infectious disease caused by a spirochete, *Treponema pallidum*, and transmitted principally by sexual intercourse, but occasionally by direct nonsexual contact, or from a pregnant woman to her fetus as in congenital syphilis. Since the introduction of penicillin, the prevalence of syphilis is declining but in no country is it a negligible health problem. *See* SPIROCHETE.

Among low economic groups living under crowded conditions there are syphilis-like diseases transmitted, usually in childhood, by nonsexual contact. These diseases have been given distinguishing names such as yaws, bejel, and pinta Other syphilis-like infections, which cannot be regarded as clinical or epidemiological entities, have been given local names such as njovera (South Rhodesia) and dichuchwa (Bechuanaland). Because of the close biological relationship of the spirochetes that cause this group of diseases including syphilis, the general term treponematosis is often used. A natural disease of rabbits, venereal spirochetosis, is closely related biologically to the human treponematoses. *See* BEJEL; PINTA, YAWS.

Treponema pallidum, cause of syphilis (General Biological Supply House, Inc.)

Causative organism. *T. pallidum* is a thin spiral organism varying in length from 5 to 20μ In fluid medium it resembles a corkscrew with 4-14 regular spirals and moves with a rapid rotary motion with little movement of translation. In a viscid medium it has a snakelike motion with elongated spirals. It is visualized with difficulty under the ordinary light microscope, but shows clearly under the darkfield microscope. Electron microscopy shows a spiral central axial filament with surrounding periplast and terminal flagella-like projections The organism stains poorly. Inoculation into the skin or testes of rabbits induces characteristic lesions. Monkeys, hamsters, guinea pigs, and mice develop infection but show lesions irregularly Pathogenic treponemes cannot be grown on artificial media or tissue culture. Virulence is maintained for 7–14 days in a special anaerobic medium, and indefinitely when frozen at approximately −76°C, the temperature of dry ice

A number of strains of treponemal-like spirochetes have been grown on artificial media under anaerobic conditions, but these are invariably nonpathogenic, their original source is uncertain, and there is only limited antigenic relationship with disease-producing treponemes

Syphilis in man. The disease evolves in a stepwise manner. The primary or initial lesion develops at the site of implantation of treponemes within 1–6 weeks. Characteristically this lesion, also referred to as the hard chancre, is indurated, often like cartilage, and the lymph node draining the area is enlarged The primary lesion is most often located on the genitalia, however, in about 10% of cases the primary lesion is extragenital, most commonly on the lips or fingers. Fluid expressed from the chancre contains *T pallidum* in abundance

Treponemes are soon widely disseminated by the lymphatics and blood giving rise, after an interval of 3–8 weeks, to generalized lesions characteristic of the secondary stage of the disease Most prominent in this stage are treponeme-containing lesions of the skin and mucous membranes of the lips, mouth, and genitalia Foci of spirochetal multiplication are present in many organs and often give rise to symptoms of arthritis, iritis, meningitis, and hepatitis Low grade fever, loss of appetite, and general malaise are common

Even without treatment lesions of the secondary stage eventually subside and the disease enters the tertiary stage, or latent phase, in which, despite the persistence of infection, no lesions may be detectable; or isolated lesions in which treponemes are scarce may develop at irregular intervals in the skin, the liver, the central nervous system, or the

heart and great vessels. Formerly, syphilis was an important cause of insanity (general paresis) and of cardiovascular disease *See* PARESIS, GENERAL. During the primary stage two types of antibodies develop One, a biologically nonspecific antibody, is detectable by complement fixation and flocculation tests such as the Wasserman, Kolmer, Kahn, Kline, and Eagle, all of which utilize an antigen made of normal beef heart. A more specific antibody is detected by the Treponemal Immobilization Test (TPI) utilizing living *T. pallidum* as antigen. Both types of tests remain positive for many years, often for life, in the absence of treatment. *See* ANTIBODY; COMPLEMENT-FIXATION TEST; SEROLOGY

Clinical or epidemiological evidence may suggest syphilis, but definitive diagnosis depends upon demonstration of *T. pallidum* by darkfield examination or upon positive serological tests. Formerly arsenicals and bismuth were the principal form of treatment, but these have been superceded by penicillin.

Prevention. Avoidance of intimate contact with an infected person is the surest preventive. Infec- [illegible]

[illegible] water destroys contaminating treponemes. Penicillin either by mouth or by injection is also an effective preventive. Public health measures should be directed to detecting and treating infected persons, particularly women, in whom lesions are often obscured. No vaccine is available for any of the treponematoses. [T.B T.]

Systems engineering

The design, prediction of performance, building, and operation of large and [illegible] tions of el [illegible] the requis [illegible] under changing conditions of load, environment, and information inputs

A system is a collection of matter within prescribed boundaries The behavior of the system, as perceived by an observer stationed outside the system, is described in terms of system quantities or variables Figure 1 is a schematic representation of a system.

A system quantity, or variable, is any characteristic of a system measurable by an observer stationed outside the system An input quantity is a variable whose value at any instant of time is determined by events occurring outside the system

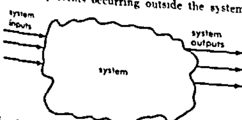


Fig 1 Schematic representation of a system.

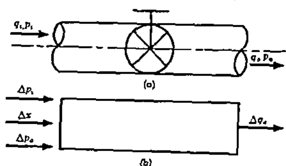


Fig 2. Valve-controlled flow system. (a) Actual system (b) Schematic representation.

An output quantity is a variable whose value at any instant of time is determined by events occurring within the boundaries of a system in response to changes in input quantities. A constant quantity is a characteristic of a system whose value does not change with time over the interval under observation.

The definitions of input and output quantities given here are highly idealized In actual practice all input values are influenced by what goes on within the system. Similarly all outputs are dependent upon what goes on outside the system. In many cases the ideal definition is a useful approximation of the actual situation. The flow device shown in Fig 2*a* is a simple system which may be schematically represented as shown in Fig 2*b*. Where the fluid is incompressible, the change in output flow is

$$\Delta q_o = f(\Delta p_o, \Delta p_i, \Delta x)$$

where Δq_o is change in output flow q_o, Δp_o is change in exit pressure p_o, Δp_i is change in entrance pressure p_i, and Δx is change in valve opening It can be shown that for small changes in the variables

$$\Delta q_o = K_1 \Delta p_o + K_2 \Delta p_i + K_3 \Delta x$$

The model of the system in Fig. 2*b* is adequate to express the behavior of the system for small changes in the inputs Δx, Δp_i, and Δp_o The values of the entrance and exit pressures depend only on conditions outside the system, providing the system is fed from a sufficiently large reservoir and exhausts into a sufficiently large reservoir so that changes in the flow Δq_o do not produce significant changes in Δp_i and Δp_o. Similarly the change in valve opening Δx is set outside the system and is not altered by changes in conditions within the system. Actually changes of pressure in the system do change the valve opening, but this influence is normally not a significant one. The terms input and output are widely used, whether they conform to the ideal definition or not.

System performance or behavior. The response of a system to changes in the inputs is called its performance or behavior Where the inputs are functions of time, the behavior of the output quantities is a function of the inputs, the system char-

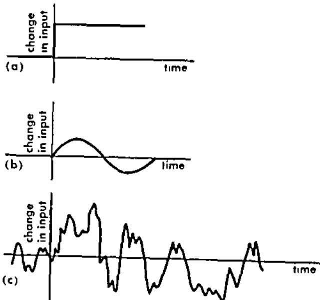


Fig 3. Types of inputs. (a) Step-change input. (b) Sinusoidal-change input. (c) Random-change input.

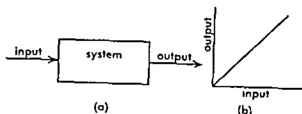
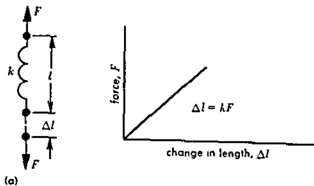
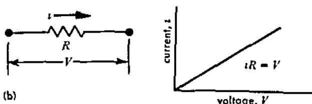


Fig. 4 A linear system. (a) Schematic representation (b) Response characteristic.





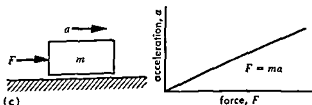


Fig. 5. Examples of linear systems (a) Spring (b) Electrical resistance. (c) Mass

acteristics, and time, and the behavior is known as the transient response. Where the inputs are not changing with time, the performance is called the steady-state response.

Standard types of input. The availability of certain analytic techniques and measurement capabilities has made the determination of system behavior convenient in terms of response to specific inputs. Such inputs include: (1) a step (indicial) change; the input is changed from one steady state value to another instantaneously (Fig 3a); (2) a steady sinusoidal input; the input is varied sinusoidally with time (Fig. 3b); (3) a random-change input; the input varies randomly with time (Fig. 3c).

Linear systems. Linear systems are those in which an effect (output) is directly proportional to its cause (input) as shown in Fig. 4

While few systems are completely linear, many exhibit substantially linear behavior, and the analysis that has been developed for such systems can be widely and successfully applied. Examples of linear systems include a spring, an electrical resistance, and a mass (see Fig. 5).

Nonlinear systems In these systems at least one element possesses nonlinear characteristics Any element with an output whose change is not proportional to the change in input is nonlinear. As a matter of fact, all systems are nonlinear, but in many instances the deviation from linearity is small enough so that linear theory accurately predicts performance.

The analysis of nonlinear systems is much more difficult than that of linear systems. There are many such systems for which no known closed generalized mathematical solution exists. Many important systems are nonlinear, for example, the automatically controlled heating system used in many homes is an example of an ON-OFF system The heat source (the furnace) is fully on if the house temperature is below a certain desired level The burner is shut off when the house temperature is above the level desired The output energy of the furnace is not proportional to the error in temperature (see Fig 6).

A system that exhibits saturation, such as an electric motor, is also nonlinear The output is proportional to the input until the magnitude of the input exceeds a certain value. Further increases in input result in no additional change in the output (see Fig. 7)

Backlash is another form of nonlinearity encountered in many systems. Upon reversal of the input, the output does not change until the input has changed by an amount equal to the backlash, or play, in the system (see Fig 8)

Many commonly used control elements exhibit what is called a dead zone. For example, some hydraulic control valves are designed to produce an output flow proportional to the input signal However, it is extremely difficult to build such a valve that will not exhibit a dead zone at its center position as shown in Fig. 9

energy input

T^* temperature

T^* = desired temperature

Fig 6 ON-OFF system.

output

input

Fig 7. Saturation

While some nonlinearities are regarded as undesirable, the deliberate use of certain nonlinear elements in systems is of vital importance. Hence, there is continuing effort to increase our ability to analyze the dynamic behavior of systems containing such elements

Systems design. The design of large-scale systems is not new. The construct[illegible] road s[illegible] has lo[illegible] involving able leadership and coordinated group-work. Nevertheless, during the last two decades a set of concepts and tools has been developing that is proving to have real power in dealing with large an[illegible] [illegible] problems will demand in the future.

Design is the bringing into being of something that did [illegible] cle[illegible] lem[illegible] are [illegible] bee[illegible] by a poor and narrow statement of what, in fact, was a broad situation. Step two is the gathering of all possible information. Goals must be outlined, conditions to be faced must be clearly stated. Each statement that places limitations on action must be carefully checked. At this point, or perhaps sooner, possible solutions begin to appear, the more different possible solutions the better. Each possible solution poses searching questions as to the statement of goals and limitations.

A relative evaluation of the possible solution must then be undertaken, and a decision must be made as to the promising one to pursue. In the process, and in the continued work on the design chosen, powerful methods are available to predict system performance, to compute the influence of changes in system quantities upon behavior, and to arrive at an optimum design. Some of the methods and areas of knowledge most widely used are briefly discussed below.

Dynamics. Many of the most difficult systems problems concern phenomena that change with time. In an increasing number of situations designs based upon steady-state conditions are no longer adequate. The complexity of modern systems makes prediction of dynamic behavior extremely difficult, but solution is possible if skillful advantage is taken of (1) the judicious selection of simplified, but reasonably accurate, models of the actual system, and (2) the proper use of computing aids.

All complex systems taken as they exist are too complicated to deal with. Hence, the selection of a model, simple enough to be expressible in mathematical terms, is a key step. In the world of physical inanimate components, a well-developed body of laws exists which expresses the behavior of sys-

output

input

Fig 8 Backlash

output

input

dead zone

Fig 9. Dead zone

tems consisting of such components. The form such expressions take, however, depends upon the viewpoint and objective of the observer of the system behavior. The gross system behavior in the presence of known disturbances of the system and of its components, not many orders different in size and relatively limited in number, may be expressed as ordinary or partial differential equations. A clear and relatively simple expression ties the behavior of every element to every other element. Thus, in an elastic structure subjected to load disturbances, the vibration of each element can be directly determined by the application of Newtonian mechanics. However, the description of the behavior of individual molecules of gas in a system comprising a great many such molecules cannot be conveniently handled in the same manner In such cases statistical mathematics and probability concepts must be used. The distinction is not one of difference in systems, but rather the behavior of the system one is interested in Thus thermodynamics and statistical mechanics are concerned with identical kinds of systems viewed with different objectives in mind. Since every model is a simplified picture of the actual system, its validity must be checked by measurements made on the actual system.

The growing size and complexity of the systems being dealt with means that even the simplest useful model often requires the use of computing machines of considerable scope. Both digital and analog machines are used.

Optimization of system performance The cost of large systems is great, requiring optimum design and operation of such systems. Often the basis for the measurement of performance is dollars. The design that produces the desired result at minimum cost often is regarded as optimum Even with this criterion, however, difficult problems arise. The load that the system must handle may not be known. The optimum design for one load level usually is not an optimum design for all operating levels It may be necessary to express the inputs and outputs of the system in statistical form. In addition, the properties of the system itself may change with time These facts have placed increasing emphasis upon the development of self-adaptive systems, in which a continuous or periodic measurement of input and output quantities is made This information is used to compare the actual behavior of the system with some prescribed optimum behavior. If such comparison reveals a difference, the system changes its properties automatically in a manner to approach more nearly the optimum design

Examples of complex systems. The engineer is called upon to face such extremely complex problems as (1) the design of a continental defense system with adequate detection capability, adequate communications, suitable arms and manpower, and a clear understanding of how important decisions are to be made; (2) the design of a transportation system involving rail, water, road, and air transportation optimized to meet the changing needs of our society; and (3) the design of an automatic, inexpensive, and more rapid national mail service. The construction of such systems is enormously expensive and involves long time commitments of a serious nature. It is essential that the solutions developed be as nearly optimum as possible The problem of optimization, however, is not easy. The exact loads that such complex systems will face are not known and often can be expressed only in statistical fashion.

The design and building of many large systems encompass many fields of endeavor, such as the various fields of engineering and science, mathematics and computation, and the social sciences. Hence, specialists in many areas must be called upon. Perhaps the greatest responsibility, however, falls upon those who carry the burden of the design of the over-all system from initial conception to a working prototype. Such leaders must have broad outlooks. While they cannot be specialists in every field, they must be able to communicate with and lead groups of specialists and bring the best tools available to bear upon the problems which arise in large systems. *See* CONTROL SYSTEMS; CYBERNETICS; HUMAN ENGINEERING.

[J A H]

Component design. By this process the physical form of the elements that comprise the system are brought into existence. In the design of a system, the design of the components of the system should be considered coincidentally with the progress of the system design Reiterative consideration of the design of the components and the design of the system is essential to the optimum design of both

A component is an element of an engineering system. It is defined in terms of, and satisfies, a functional requirement. Usually a component is sufficiently well-defined physically that it is easily distinguishable from the other components in the system. For example, it may be an electric motor, a hydraulic servovalve, or an electronic amplifier Components may also be combined with other components and packaged together. In systems where weight or volume are at high premium, as in missiles, the components are so tightly integrated physically one with another that it may be difficult to identify specific components.

Functional specifications. The basic system expressed in block diagram form defines the functional requirements imposed upon the components, establishing the component specifications. Component design converts these functional specifications to drawings and specifications, which completely define the component form, materials, construction techniques, finishes, and testing procedures, thus making possible the manufacture of the component

The process of design is difficult to define, but one can roughly outline the successive stages through which a component design proceeds. The initial stage is the most elusive one. The functional specifications prescribe s[illegible] 'ationship between

'a input and an output variable parameter. The 3le of the design engineer is to determine the physical form of the optimum device for accomplishing this goal. With an awareness of physical principles and a flare for the innovative and ingen- he considers a variety of ways in which this nay be satisfied. He then analyzes and studies ntative and often graphical descriptions of everal approaches, applying his knowledge of physical sciences, his engineering experience the advantages and disadvantages of the vari- means, and his knowledge of the techniques economics of manufacturing processes in- d From this many-dimensioned comparison, or more approaches that seem best suited to articular problem are selected

eas thus conceived and analyzed are reduced lesign draftsmen to drawing descriptions ade- e for the fabrication of the components in an rimental shop. The experimental hardware is iected to the scrutiny of experimental testing isted, and changed in accordance with these s and new drawing descriptions are made of component to satisfy better the functional spec- tions The design is ultimately defined in an mbly drawing which shows the relationships ong all parts of the *component*, and a number detail drawings each of which completely de- bes a part See ENGINEERING DRAWING

n addition to satisfying the functional specifi- tions as given in the block diagram, the engineer- t designer must consider everything which may ect the performance of the component. These her considerations also play a vital role in the itial innovative stage of the design process

Environmental design One such consideration is e recognition of the environment in which the mponent *must perform* its functional task En- ronmental requirements can range from the con- olled temperature pressure, and humidity sur- ndings of a chemical process-plant control stem to the wildly fluctuating uncontrolled envi- onment of a missile control system The design ngineer must recognize the *influence* of the envi- ronment on material *choices* (temperature in- fluences on physical properties) on dynamic oper- ation (high accelerations may cause high stresses or vibrations), and on other effects See ENVIRON- MENTAL TEST.

Reliability If a system is to perform properly, its components must be reliable Therefore, the probability of failure of individual components cannot be *overlooked* The reliability of a compo- nent its ability to perform its intended function at the required time, is strongly related to initial de- sign *For* example all design analyses must be based on conservative assumptions, and the design engineer *must* anticipate potential weak spots Simplicity of design is important, because a greater number of parts in a component means a greater probability of component failure due to the failure of a single part See RELIABILITY OF EQUIPMENT.

Maintainability Since 100% reliability is unattainable, another important consideration in component design is provision for maintenance. Depending upon system requirements, such maintenance may range from component replacement upon any indication of unsatisfactory performance, in which case design emphasis will be on identification of faulty performance and easy removal, to on-the-job repair of the component where design emphasis will be on accessibility. *See* MAINTAINABILITY OF EQUIPMENT.

Packaging. Along with the foregoing considerations, the design engineer must arrange the elements of the component into mutual relationships which best satisfy all the conflicting requirements. This somewhat misnamed "packaging" process must optimize the three-dimensional *arrangement* of the component elements, recognizing the component's spatial relationships to other *components* in the system, while simultaneously satisfying functional, environmental, reliability, maintenance, and other considerations. *See* MINIATURIZATION OF EQUIPMENT; PACKAGING OF EQUIPMENT.

Component testing. The design process inevitably involves the reevaluation of drawing descriptions of component conceptions as a result of experimental evaluation of the merits of such conceptions This design-test process is called development Successfully negotiated, it converges upon a component design that best satisfies the many requirements and is then *embodied in a set* of manufacturing drawings. Manufacturing drawings are subsequently transformed into physical hardware, and then the resulting production components must undergo production testing before they themselves can be integrated into the engineering system

In development test a component first must be *tested to determine* whether it satisfies adequately the functional specifications. By *means of* an experimental set-up in which the component *is an element, the* input output characteristics of the component are evaluated Some of the elements of the component may be *capable of* a certain range of adjustments; these adjustments will be systematically *altered so as to* optimize the functional performance of the component The record of optimum adjustment *positions provides* a calibration of the component, by which it is subsequently possible to achieve rapidly the desired functional performance without recourse *to extensive testing.*

The component must also be subjected to a wide *range* of environmental conditions which may influence *its functional* performance Ideally, the test is *arranged so that* the functional performance of the component may be *critically evaluated* as the component is simultaneously subjected to one or more of the environmental conditions.

Since it is not usually *possible to* define specifically and absolutely the conditions and *requirements to be* ultimately imposed upon the system, it is therefore not ...

components must satisfy. The designer of the equipment and the user jointly prepare specific tests to be performed on each component. Required performance of the component is also specified. These tests, known as qualification tests, ensure mutual understanding of the functional and environmental requirements which the component must satisfy. Components that perform satisfactorily in qualification tests are accepted for integration into the system. *See* QUALIFICATION TEST.

Subsystem test. In a complex system, combinations of components, called subsystems, are tested independently of the over-all system in order that their performance may be ascertained before the arrangement of the over-all system. It must be recognized that the mathematical models, which describe the functional specifications imposed upon the components or subsystems, are considerable idealizations of the performance of the physical components or subsystems. Therefore, it is entirely possible that while an individual component *may satisfy the functional requirements imposed* upon it, interrelationships and the crosscouplings between undefined characteristics of several components may be such that the physical combination of the components will result in operation different from that predicted from the simple combination of the component's functional specifications Subsystem testing of several components can thus identify mutual characteristics which are not ascertainable from the tests of the individual components themselves.

System test. Following a sequence of component and subsystem testing, the entire system is assembled and subjected to a series of adjustments and calibrations as defined above in component testing. The system is then tested to demonstrate satisfactory functional performance under the pertinent range of environmental conditions Again the system developer and the customer carefully prepare a statement of the functional and environmental requirements of the system in the form of system-qualifications requirements When the system satisfies these qualification tests, it is completely developed and is accepted by the customer [R W M]

Bibliography: H. H. Goode and R. E Machol, *System Engineering*, 1957.

T Tabulata to Tourmaline

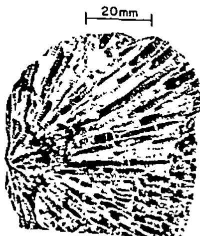
Tabulata

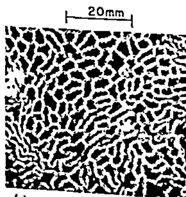
An extinct group (order) of entirely colonial corals which were numerous in the Paleozoic seas. [illegible]

[illegible] of vertical tubes which may be cylindrical, laterally compressed, or polygonal The cylindrical forms may grow erect with connecting tubules between cylinders, or they may grow attached to a mollusk or brachiopod shell. *Syringopora*, the "organ pipe" coral, is characteristic of the cylindrical forms The polygonal forms typically have rows of pores in the walls connecting the corallites *Favosites*, the "honeycomb" coral, is an example of this type A corallum composed of laterally compressed corallites which bud end to end results in the peculiar "chain coral," *Halysites*.

The major internal structure of almost all tabulate corals is a vertical series of horizontal plates called the tabulae Septa, the vertical partitions so common in the Rugose corals, are poorly developed or absent in the Tabulata In some forms twelve rows of weakly developed septal ridges or spines are present on the walls of the corallites. *See* COELENTERATA FOSSILS; RUGOSA [E.C ST.]







Tabulate corals (a) Syringopora, showing cylindrical stems and lateral connections. (b) Favosites, showing polygonal corallites, mural pores, and tabulae (c) *Halysites*, showing chain structure [illegible]

Tacan

A polar coordinate navigation system in which a single emission provides both bearing and distance information. Its effective range is 200 nautical miles. Its accuracy is ±0 2 miles in distance and 1 degree in bearing

The Tacan ground beacon operates at a frequency of from 962 to 1024 and 1151 to 1213 megacycles (Mc), where it emits pulse pairs Each pulse has a duration of about 3½ microseconds and the time between pulses of a pair is 12 microseconds Transmission occurs on 1-Mc channels, a total of 126 ground-to-air channels is provided. A total of 2700 pairs of pulses is emitted randomly. In addition, another 900 pairs of pulses are transmitted to serve as directional references. The emitted pulses are radiated by means of a nondirectional antenna around which two cylinders carrying parasitic elements rotate at a speed of 900 rpm. The inner cylinder carries a single parasitic element while the outer cylinder carries nine parasitic elements. In space, there is generated a cardioid pattern distorted with nine lobes (see Fig 2)

The Tacan receiver output consists of pulses having amplitude variations corresponding to a 15-cycle wave with a pronounced ninth harmonic. When the 135-cycle harmonic wave is extracted and compared with the pulses marking the 40° sectors, a highly precise indication of bearing is

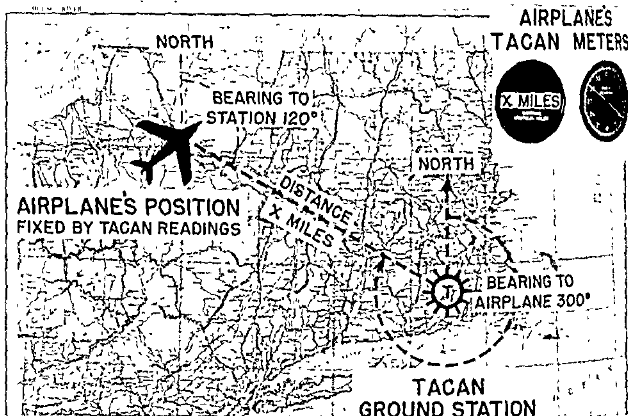


Fig. 1. Tacan provides navigational information in the form of bearing and distance. (ITT Laboratories)

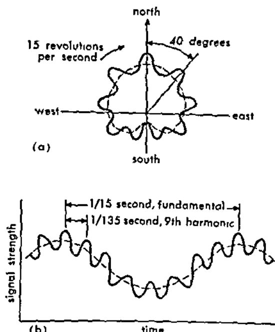


Fig. 2. Tacan (a) Antenna pattern (b) Receiver output

obtained, which is ambiguous, however, in that the particular 40° sector to which it applies is not known The extracted 15-cycle signal provides the indication necessary to resolve the ambiguity Bearing is displayed on a meter which automatically indicates bearing without ambiguity.

The Tacan airborne equipment also contains a pulse transmitter which continuously transmits pulse pairs on 126 channels in the band of from 1025 to 1150 Mc. These pulses are transmitted at a rate of about 150 per second when search in time for the reply is in process, or 30 pairs per second when the reply has been received and is being tracked These pulses are received at the ground station and retransmitted via the equipment previously mentioned. Pulses sent out in response to an airborne interrogation replace random pulses so that the total number transmitted is constant. When the pulses are returned to the receiver, the total time of transit is measured and an automatic circuit converts time to slant distance which is then indicated on a standard aircraft instrument At intervals, the pulses sent out are regularized at a frequency of 2700 cycles and keyed with the Morse code to identify the station. See NAVIGATION SYSTEMS, ELECTRONIC [PCS]

Bibliography: R I Colin and S. H. Dodington, Principles of Tacan, *Elec. Commun.*, 33:11-25, 1956; P. C Sandretto, *Electronic Avigation Engineering*, 1958.

Tachometer

An instrument that measures angular speed, as that of a rotating shaft The measurement may be in revolutions over an independently measured time interval, as in a revolution counter, or it may be directly in revolutions per minute. The instrument may also indicate the average speed over a time interval or the instantaneous speed Tachometers are used for direct measurement of angular speed and as elements of control systems to furnish a signal as a function of angular speed

Revolution counter. The simplest form of tachometer is the revolution counter shown in Fig 1 Held in contact with the r[illegible] shaft, it counts

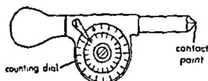


Fig 1. Revolution counter (From D M Considine, ed , *Process Instruments and Controls Handbook*, McGraw-Hill, 1957)

the number of shaft rotations It is used in conjunction with a timer to obtain average speed over the elapsed period of time Accuracy is best for uniform speeds measured over long time periods.

Chronometric tachometer. This tachometer counts the revolutions over a fixed interval of time and presents the measurement directly in terms of speed The result is the average speed for the time interval. The shorter the time interval, the more nearly the instantaneous speed will be approached.

Centrifugal tachometer. The *centrifugal force* developed by a rotating mass is a function of speed Figure 2 shows the essential parts of such a meter This form indicates instantaneous speed It is capable of accuracies on the order of ±1% of full scale value

Vibrating-reed tachometer. The vibrating reed tachometer consists of a group of reeds of different length The lowest natural frequency of vibration of each reed is a function of its length, mass and cross-sectional dimensions Observation of the reed that is vibrating forms the means of measuring the frequency of vibration (Fig. 3) The instrument is brought into mechanical contact with the device whose mechanical frequency, oscillation or rotation is to be measured by touching it to the bearing support, case, or frame The reeds may be adjusted to an accuracy of ±0 3% and are usually guaranteed to ±0 5%.

Impulse tachometer. Tachometers falling into this group may be classified as capacitor charging-current type, inductor type, or interrupted dc type

Capacitor charging-current tachometer In the instrument shown in Fig 4 the charging current of a capacitor is utilized The pickup head usually contains a reversing switch, operated from a spindle, which reverses twice with each revolution The indicator responds to the average value of these pulses The indication is proportional to time rate of these pulses and therefore to the time rate of the spindle revolutions The battery voltage must be steady and the circuit must be adjusted to the actual value of this voltage With a steady voltage, this device is capable of good accuracy

Inductor tachometer With this instrument the rotating member, which could be a piece of soft iron or laminated iron, causes the magnetic flux of a circuit containing a magnet and pickup coil, to rise and fall The rise in flux produces a pulse of one polarity and the fall of flux produces a pulse of the opposite polarity This is then rectified for a permanent-magnet movable-coil instrument The pulses may be produced by a lump of

material passing close to a coil having a permanent-magnet core. Figure 5 shows a form of this tachometer in which the soft-iron piece, or magnet, rotates close to the pickup coil.

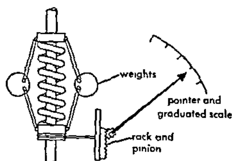


Fig 2. Centrifugal-force tachometer. (From D. M Considine, ed , *Process Instruments and Controls Handbook*, McGraw-Hill, 1957)

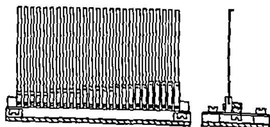


Fig 3. Vibrating-reed tachometer (From D. M Considine, ed., *Process Instruments and Controls Handbook*, McGraw-Hill, 1957)

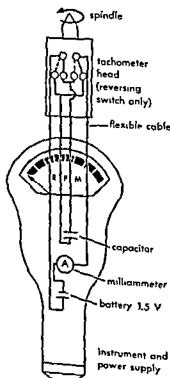


Fig 4 Capacitor-type impulse tachometer (From D M. Considine, ed , *Process Instru*

Interrupted dc tachometer. The interrupted dc of an ignition-circuit primary, whether battery-excited or magneto-excited, provides a frequency of pulses which can be used to measure speed. Figure 6 shows a frequency responsive circuit, in which current from the battery is interrupted by the contactor and excites the ignition coil. The voltage drops from these pulses excite the saturable-core transformer to produce current, which is rectified for the average-reading dc instrument.

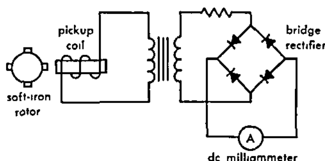


Fig 5. Schematic circuit for an inductor form of tachometer.

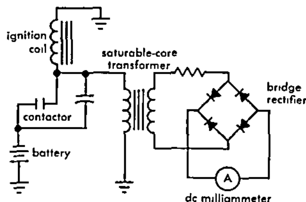


Fig 6. Schematic circuit of an interrupted dc pulse type of tachometer, as used with the ignition circuit of internal combustion engine.

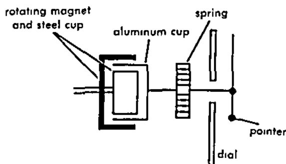


Fig. 7. Drag-type eddy-current tachometer. (*From D. M. Considine, ed., Process Instruments and Controls Handbook, McGraw-Hill, 1957*)

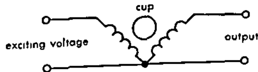


Fig 8. Schematic circuit and components of a drag-cup tachometer generator.

Eddy-current tachometer. This form, also known as drag type, is widely used for automobile speedometers and for measuring aircraft-engine speed. It contains a permanent magnet, rotated by the shaft whose speed is to be measured. Close to the revolving magnet is an aluminum disk or cup, mounted to a staff with a pointer, pivoted and free to turn against a spring (Fig. 7). The pointer is associated with a calibrated scale. As the permanent magnet revolves, eddy currents are produced in the disk or cup. The magnetic fields caused by these eddy currents produce a torque, which acts in a direction to resist the turning magnet field. The cup or disk will then turn against the spring. The disk or cup turns in the direction of the rotating-magnet field and will turn (or be dragged) until the torque developed by the eddy currents equals that of the spring. An accuracy of 10 rpm in a full scale of 3000 rpm may be obtained. The movable member may be revolved through as much as 1080°, affording high resolution in indication.

Drag-cup generators, which also use eddy currents, find a use in control systems. They have two stationary windings, positioned so as to have zero coupling, and a nonmagnetic metal cup, which is revolved by the source whose speed is to be measured (Fig. 8). One of these windings is used for excitation, inducing eddy currents in the rotating cup. These eddy currents in the cup produce a field which induces in the other winding an emf proportional to the speed of the rotating cup and at the same frequency as the exciting source. These generators usually have a high degree of linearity, low electrical noise and low starting and running torque. The output emf and energy are also low.

Velocity-head tachometer. With this type of tachometer the device whose speed is to be measured drives a pump or blower, producing a fluid flow, which is converted to a pressure. Figure shows a form of this tachometer that not only indicates the speed but produces a tape recording. This form is used on railroad locomotives. The tape is moved by the drive shaft whose speed is being measured.

Electronic tachometer. Circuits including vacuum tubes or transistors are sometimes needed to amplify weak pulses or to shape pulse waves. These pulses may be produced by photoelectric, magnetic or inductor transducers. These shaped waves are applied to frequency-responsive circuits, rectified, and applied to a dc milliammeter.

Electronic counters, also known as EPUT (events per unit of time) meters, are designed to measure frequency, whether sine-wave or pulse. The input pulse waves, clipped and shaped by discriminating circuit, are allowed to pass through a gating circuit to the displaying decade counters. The open time of the gate is controlled by a crystal oscillator or other suitable time base. The displaying counters are responsive to the number of pulses passed by the gate in a definite time. The accuracy,

established by the time-base generator, can be better than 0.01%. Since the display counters or readout may be read to ±1 count, the number of pulses applied to the input should be sufficiently high to realize its inherent accuracy. See FREQUENCY COUNTER

Electric-generator tachometer. A widely used and flexible form of electric tachometer comprises a combination of electric generator and indicator. There are two principal forms of these tachometers, (1) a dc generator with a dc voltmeter and (2) an ac generator with an ac voltmeter (or dc voltmeter with rectifier). In either case, the emf developed is proportional to the shaft speed

The ac generator form may include a circuit that is responsive to frequency but is affected only slightly by voltage, giving greater accuracy.

DC generator The dc tachometer is a small permanent magnet generator with an output of 2–10 volts per 1000 rpm A high resistance voltmeter, calibrated in rpm, indicates the speed

The dc generator assumes a polarity dependent upon direction of rotation. When used with a zero-center indicator, the instrument will indicate the direction of rotation With standardizing, an accu[ra]cy in the order of 0.25–0.1% may be realized [Se]e DIRECT-CURRENT GENERATOR

The low starting and running torque makes it [us]eful for measuring wind velocity In addition to [me]asuring speed, the dc tachometer is used as a [st]abilizing component in velocity servomechanisms. [Se]e CONTROL SYSTEMS

AC generator The ac tachometer can be con[st]ructed with a stationary winding and a revolving [pe]rmanent magnet field. Both generated voltage [an]d frequency are proportional to speed of rotation The voltage may be rectified and applied to a [p]ermanent magnet moving coil instrument calibrated in rpm. See ALTERNATING-CURRENT GENERATOR

By the addition of a saturable-core transformer [to] the instrument circuit, as shown in Fig 10, the [in]strument indication becomes frequency-respon[si]ve and only slightly affected by voltage, affording [g]reater accuracy.

The difference of two speeds may be measured [and] indicated by connecting the outputs of two [s]peed measuring circuits to a differential bridge as [s]hown in Fig 11 The individual speeds may also [b]e indicated This difference of speed indication is [i]ndependent of the actual speeds

If one of the generators is replaced by a fixed frequency, the other speed may be measured in reference to this frequency or the scale may be calibrated in terms of the speed This circuit also affords the means for suppressing as much as 90% of a speed range and allowing the top 10% to occupy the entire scale [A.H.WO.]

Bibliography: M F. Behar (ed), *Handbook of Measurement and Control*, 1951, D M Considine (ed.), *Process Instruments and Controls Handbook*, 1957; E. A Griffiths, *Engineering Instruments and Meters*, 1920

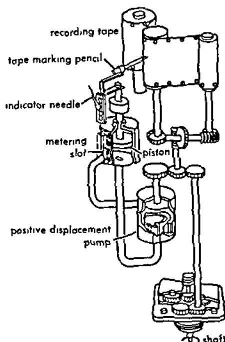


Fig 9. Sectional view of velocity-head-type tachometer (From D M Considine, ed, *Process Instruments and Controls Handbook*, McGraw-Hill, 1957)

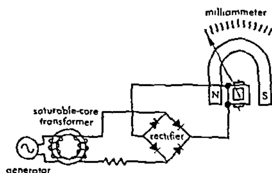


Fig 10 Schematic circuit of an ac tachometer in which a saturable-core transformer is used. (From D M Considine, ed, *Process Instruments and Controls Handbook*, McGraw-Hill, 1957)

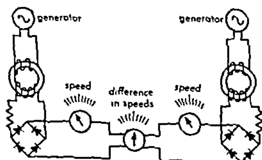


Fig 11. System employing two ac tachometers for indicating individual and differential speeds. (From D M Considine, ed, *Process Instruments and Controls Handbook*, McGraw-Hill, 1957)

Interrupted dc tachometer. The interrupted dc of an ignition-circuit primary, whether battery-excited or magneto-excited, provides a frequency of pulses which can be used to measure speed. Figure 6 shows a frequency responsive circuit, in which current from the battery is interrupted by the contactor and excites the ignition coil. The voltage drops from these pulses excite the saturable-core transformer to produce current, which is rectified for the average-reading dc instrument.

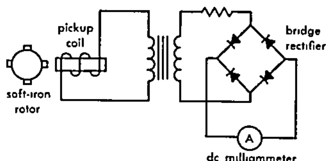


Fig. 5. Schematic circuit for an inductor form of tachometer.

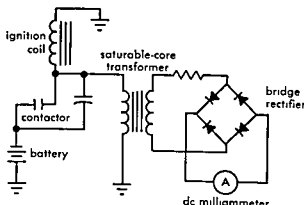


Fig 6. Schematic circuit of an interrupted dc pulse type of tachometer, as used with the ignition circuit of internal combustion engine.

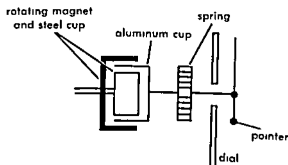


Fig 7. Drag-type eddy-current tachometer (From D. M. Considine, ed, *Process Instruments and Controls Handbook*, McGraw-Hill, 1957)

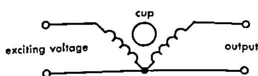


Fig. 8. Schematic circuit and components of a drag-cup tachometer generator.

Eddy-current tachometer. This form, also known as drag type, is widely used for automobile speedometers and for measuring aircraft engine speed It contains a permanent magnet, rotated by the shaft whose speed is to be measured Close to the revolving magnet is an aluminum disk or cup mounted to a staff with a pointer, pivoted and free to turn against a spring (Fig. 7). The pointer is associated with a calibrated scale. As the permanent magnet revolves, eddy currents are produced in the disk or cup. The magnetic fields caused by these eddy currents produce a torque, which acts in a direction to resist the turning magnet field. The cup or disk will then turn against the spring. The disk or cup turns in the direction of the rotating-magnet field and will turn (or be dragged) until the torque developed by the eddy currents equals that of the spring. An accuracy of 10 rpm in a full scale of 3000 rpm may be obtained. The movable member may be revolved through as much as 1080°, affording high resolution in indication.

Drag-cup generators, which also use eddy currents, find a use in control systems. They have two stationary windings, positioned so as to have zero coupling, and a nonmagnetic metal cup, which is revolved by the source whose speed is to be measured (Fig. 8). One of these windings is used for excitation, inducing eddy currents in the rotating cup. These eddy currents in the cup produce a field which induces in the other winding an emf proportional to the speed of the rotating cup and at the same frequency as the exciting source. These generators usually have a high degree of linearity, low electrical noise and low starting and running torque. The output emf and energy are also low.

Velocity-head tachometer. With this type of tachometer the device whose speed is to be measured drives a pump or blower, producing a fluid flow, which is converted to a pressure. Figure 9 shows a form of this tachometer that not only indicates the speed but produces a tape recording. This form is used on railroad locomotives The tape is moved by the drive shaft whose speed is being measured

Electronic tachometer. Circuits including vacuum tubes or transistors are sometimes needed to amplify weak pulses or to shape pulse waves These pulses may be produced by photoelectric, magnetic or inductor transducers. These shaped waves are applied to frequency-responsive circuits, rectified, and applied to a dc milliammeter.

Electronic counters, also known as EPUT (events per unit of time) meters, are designed to measure frequency, whether sine-wave or pulse. The input pulse waves, clipped and shaped by a discriminating circuit, are allowed to pass through a gating circuit to the displaying decade counters. The open time of the gate is controlled by a crystal oscillator or other suitable time base. The displaying counters are responsive to the number of pulses ... acy, passed by the gat ... ite time.

layers bonded together by weak, secondary valencies. The mineral is therefore extremely soft (hardness 1 on Mohs scale) and possesses a perfect basal cleavage. The atoms within the layers are strongly bonded so that the mineral is highly stable, both to acids and to thermal treatment. See SILICATE MINERALS, SOLID-STATE CHEMISTRY.

The word talc or soapstone is applied also to massive materials which may contain many other mineral constituents than talc itself. The more common associated minerals are other magnesian silicates such as serpentine, chlorite, tremolite, and the carbonate minerals, calcite, dolomite, and magnesite. Steatite is a term applied to relatively pure [illegible] and 4% Al_2O_3. See SOAPSTONE.

Talc occurs as a secondary mineral resulting from the hydration of magnesium-bearing rocks and the alteration of minerals such as pyroxenes, amphiboles, and olivine. Fibrous talc is often closely associated with, and probably derived from, tremolite, and the fibrous character is inherited from the parent mineral. The major talc-producing states are New York, California and North Carolina. The major talc producing areas of the world are the United States, France, Italy, Austria, and Japan (precise data for the U.S.S.R. and China are not available).

Talc is a widely used raw material. Figures for talc sold or used by the producers in the United States in 1957 show that over 30% went into ceramic applications, 20% was used in paints, and about 10% for insecticides. The remainder was used in a wide variety of applications. In the ceramic industry talc is used in many whiteware bodies, tableware, electrical porcelain, high frequency insulation, and glazed wall tiles. Talc apparently imparts greater resistance to mechanical stresses arising from temperature differences, and may also prevent crazing. In paints, talc is used as an extender and as a pigment. It is used as a filler for paper, rubber, and asphalt. In the cosmetics industry, it is used in toilet powders, soaps, and creams. Massive talc is cut into slabs and used for laboratory tables, sinks, sanitary appliances, acid tanks, electrical switchboards, mantels, and hearthstones.

[G. W. BR.]

Tall oil

A by-product from the sulfate process for making cellulose. Tall oil and its derived products find industrial applications in the production of adhesives, binders, wetting agents, soaps, driers, emulsifiers, flotation agents, linoleum, printing inks, textile finishes, and varnishes.

Pine wood is digested under pressure with alkali and sodium sulfate. Rosin acids and fatty acids are extracted from the pulp as soaps. So-called sulfate turpentine is also recovered during the treatment. Addition of mineral acids to the soaps of the rosin acids and fatty acids, and subsequent refining gives a mixture of rosin and fatty acids known as liquid rosin, or tall oil. One ton of pulp yields as much as 100 lb of tall oil. Over 350 000 tons of crude tall oil was produced in 1958 in the United States. Tall oil yields rosin and fatty acids at lower cost than other sources.

Refining of tall oil by fractionation began on an extensive scale in 1942. Details vary from plant to plant, but all basically use injected steam and reduced pressures to keep rosin and fatty acids from decomposing. The crude rosin and fatty acids are further up-graded to fulfill specific requirements.

About 90% of tall oil is composed of acidic materials, containing approximately equal parts of rosin acids and fatty acids. The remainder is a complex mixture of fatty acid esters, sterols, higher alcohols, hydrocarbons, and decomposition products resulting from the prolonged high temperatures used to distill the oil. The fatty acids are mainly oleic and linoleic acids. Smaller amounts of linolenic and palmitic acids are also present.

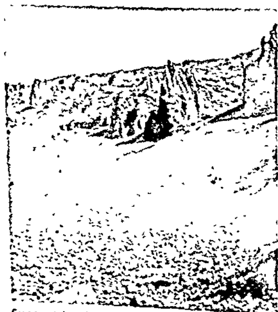
Modern methods of refining have produced grades of rosin and fatty acids each containing as little as 3% of the other. Treatments have been developed to remove odor and color. Purification of tall oil by partition between immiscible solvents has been explored. If tall oil is cooled, rosin acids will crystallize. Recrystallization from suitable solvents such as methanol gives pure rosin acids.

Attempts are being made to develop tall oil rosin as a possible substitute for gum and wood rosin. Future research will determine the feasibility of this replacement. See DRYING OIL; ROSIN, TURPENTINE, WOOD CHEMICALS

[E. L. S.]

Talus

A heap or sloping sheet of loose rock waste at the base of a cliff or steep slope. Talus and its English equivalent, scree, are terms properly applied to the



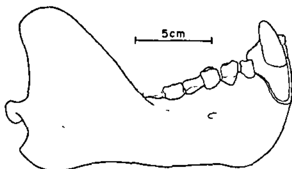
Fragmental rocks on north slope of Mud Creek Canyon, above the timber lines, Siskiyou County, California (USGS)

Taconite

The name given to the siliceous iron formation from which the high-grade iron ores of the Lake Superior district have been derived It consists chiefly of fine-grained silica mixed with magnetite and hematite. As the richer iron ores approach exhaustion in the United States, taconite becomes more important as a source of iron. To recover the ore mineral in a usable form for the production of iron, taconite must be finely ground, and the magnetite or hematite concentrated by a magnetic or other process. Finally, the concentrate must be agglomerated into chunks of size and strength suitable for the blast furnace. *See* IRON (EXTRACTION FROM ORE); ORE AND MINERAL DEPOSITS; ORE DRESSING. [C S HU]

Taeniodonta

These extinct quadrupedal land mammals are known from early Tertiary deposits in the United States. One group, the Stylinodontinae, are adaptively convergent with many of the ground sloths; for in their evolutionary succession is seen the development of large size, of deep, massive lower jaws, of a reduced number of simplified, peglike upper and lower teeth with enamel remaining only as lateral bands, of curved and somewhat shearing



Side view of lower jaw of a middle Paleocene taeniodont, *Psittachotherium* (After Matthew, 1937)

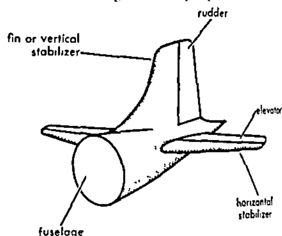
canines resembling rodent incisors, and of large, laterally-flattened claws. This succession leads from *Wortmania* of the early Paleocene to *Stylinodon* of the upper Eocene The animals of the other group, Conoryctinae, ranging from *Onychodectes* of the early Paleocene to *Conoryctes* of the middle Paleocene, remain small to medium-sized and have less specialization in dentition and skeleton. Conoryctines developed enlarged canines, but the lower jaws were unspecialized and the cheek teeth, though simplified, remained low-crowned, enamel-enclosed, and cuspidate.

Taeniodonts probably arose from a generalized Cretaceous stock that would presently be assigned to the Insectivora. *See* INSECTIVORA FOSSILS. [D E.S.]

Tail assembly

The assembly of the vertical tail, the horizontal tail, and a small section of the rear of the fuselage, as illustrated

The vertical tail consists of a fin which is fixed to the fuselage or body and a rudder which is movable by the pilot. The fin (fixed portion) is a symmetrical airfoil which is in line with the center line of the fuselage. In steady flight, the rudder is



Principal parts of tail assembly.

stationary and approximately in line with the fin The purpose of the vertical tail is to keep the air plane in line with the direction of motion or relative wind velocity.

In actual practice, the rudder may be slightly offset from the fuselage centerline to produce a small force on the fin to balance any slight difference in drag of the wings

The span of the horizontal tail lies in a horizontal plane and usually consists of a stabilizer and elevator The stabilizer and elevator are symmetrical airfoil sections whose usual position is not in line with the fuselage centerline.

The principal purpose of the horizontal tail is to provide a balancing tail load so that the weight of the airplane. the lift. and the horizontal tail load are in equilibrium for a specific airspeed. *See* AIRPLANE.

The elevator is moved by the motion of the pilot's control stick and can be moved rapidly whereas the stabilizer can be moved slowly by means of a trim control which is independent of the control stick motion.

Airplanes which operate at speeds greater than 80% of the speed of sound have tail surfaces which are sweptback 35–40°. *See* STABILIZER

Airplanes which operate at supersonic speeds usually have the horizontal tail sweptback and in one piece which is movable and is controlled by the motion of the pilot's control stick. Such a surface is often called a stabilator. *See* AIRFRAME. [R G BO

Talc

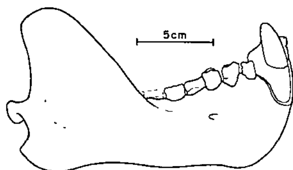
A hydrated magnesium silicate approximating in composition to $Mg_3Si_4O_{10}(OH)_2$. Chemical analyses frequently show percentages of calcium oxide, CaO, aluminum oxide, Al_2O_3, and other oxides, but these arise mainly from impurities. Small amounts of Al_2O_3 may enter the talc structure The structure is closely related to that of the micas and consists of electrically neutral magnesium silicat

Taconite

The name given to the siliceous iron formation from which the high-grade iron ores of the Lake Superior district have been derived. It consists chiefly of fine-grained silica mixed with magnetite and hematite. As the richer iron ores approach exhaustion in the United States, taconite becomes more important as a source of iron. To recover the ore mineral in a usable form for the production of iron, taconite must be finely ground, and the magnetite or hematite concentrated by a magnetic or other process. Finally, the concentrate must be agglomerated into chunks of size and strength suitable for the blast furnace. *See* IRON (EXTRACTION FROM ORE); ORE AND MINERAL DEPOSITS; ORE DRESSING. [C.S.HU.]

Taeniodonta

These extinct quadrupedal land mammals are known from early Tertiary deposits in the United States. One group, the Stylinodontinae, are adaptively convergent with many of the ground sloths; for in their evolutionary succession is seen the development of large size, of deep, massive lower jaws, of a reduced number of simplified, peglike upper and lower teeth with enamel remaining only as lateral bands, of curved and somewhat shearing



Side view of lower jaw of a middle Paleocene taeniodont, *Psittacotherium* (After Matthew, 1937)

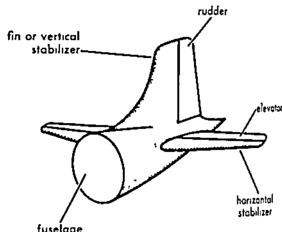
canines resembling rodent incisors, and of large, laterally-flattened claws. This succession leads from *Wortmania* of the early Paleocene to *Stylinodon* of the upper Eocene. The animals of the other group, Conoryctinae, ranging from *Onychodectes* of the early Paleocene to *Conoryctes* of the middle Paleocene, remain small to medium-sized and have less specialization in dentition and skeleton. Conoryctines developed enlarged canines, but the lower jaws were unspecialized and the cheek teeth, though simplified, remained low-crowned, enamel-enclosed, and cuspidate.

Taeniodonts probably arose from a generalized Cretaceous stock that would presently be assigned to the Insectivora. *See* INSECTIVORA FOSSILS. [D.E.S.]

Tail assembly

The assembly of the vertical tail, the horizontal tail, and a small section of the rear of the fuselage, as illustrated.

The vertical tail consists of a fin which is fixed to the fuselage or body and a rudder which is movable by the pilot. The fin (fixed portion) is a symmetrical airfoil which is in line with the center line of the fuselage. In steady flight, the rudder is



Principal parts of tail assembly.

stationary and approximately in line with the fin. The purpose of the vertical tail is to keep the airplane in line with the direction of motion or relative wind velocity.

In actual practice, the rudder may be slightly offset from the fuselage centerline to produce a small force on the fin to balance any slight difference in drag of the wings.

The span of the horizontal tail lies in a horizontal plane and usually consists of a stabilizer and elevator. The stabilizer and elevator are symmetrical airfoil sections whose usual position is not in line with the fuselage centerline.

The principal purpose of the horizontal tail is to *provide a balancing tail* load so that the weight of the airplane, the lift, and the horizontal tail load are in equilibrium for a specific airspeed. *See* AIRPLANE.

The elevator is moved by the motion of the pilot's control stick and can be moved rapidly, whereas the stabilizer can be moved slowly by means of a trim control which is independent of the control stick motion.

Airplanes which operate at speeds greater than 80% of the speed of sound have tail surfaces which are sweptback 35–40°. *See* STABILIZER.

Airplanes which operate at supersonic speeds usually have the horizontal tail sweptback and in one piece which is movable and is controlled by the motion of the pilot's control stick. Such a surface is often called a stabilator. *See* AIRFRAME. [R.G.BO.]

Talc

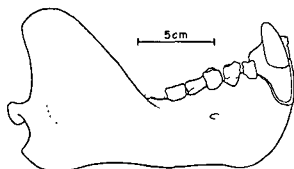
A hy ... in
comp
yses
CaO, aluminum oxide, Al_2O_3, and other oxides, but these arise mainly from impurities. Small amounts of Al_2O_3 may enter the talc structure. The structure is closely related to that of the micas, and consists of electric ... tral magnesium silicate

Taconite

The name given to the siliceous iron formation from which the high-grade iron ores of the Lake Superior district have been derived. It consists chiefly of fine-grained silica mixed with magnetite and hematite As the richer iron ores approach exhaustion in the United States, taconite becomes more important as a source of iron. To recover the ore mineral in a usable form for the production of iron, taconite must be finely ground, and the magnetite or hematite concentrated by a magnetic or other process. Finally, the concentrate must be agglomerated into chunks of size and strength suitable for the blast furnace. *See* IRON (EXTRACTION FROM ORE); ORE AND MINERAL DEPOSITS; ORE DRESSING. [C S.HU.]

Taeniodonta

These extinct quadrupedal land mammals are known from early Tertiary deposits in the United States. One group, the Stylinodontinae, are adaptively convergent with many of the ground sloths; for in their evolutionary succession is seen the development of large size, of deep, massive lower jaws, of a reduced number of simplified peglike upper and lower teeth with enamel remaining only as lateral bands, of curved and somewhat shearing



Side view of lower jaw of a middle Paleocene taeniodont, *Psittachotherium* (After Matthew, 1937)

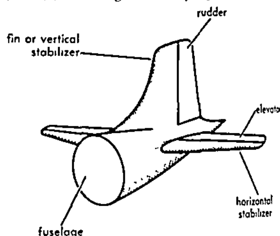
canines resembling rodent incisors, and of large, laterally-flattened claws. This succession leads from *Wortmania* of the early Paleocene to *Stylinodon* of the upper Eocene. The animals of the other group Conoryctinae, ranging from *Onychodectes* of the early Paleocene to *Conoryctes* of the middle Paleocene, remain small to medium-sized and have less specialization in dentition and skeleton Conoryctines developed enlarged canines, but the lower jaws were unspecialized and the cheek teeth though simplified, remained low-crowned, enamel-enclosed, and cuspidate.

Taeniodonts probably arose from a generalized Cretaceous stock that would presently be assigned to the Insectivora. *See* INSECTIVORA FOSSILS. [D F.S]

Tail assembly

The assembly of the vertical tail, the horizontal tail, and a small section of the rear of the fuselage, as illustrated.

The vertical tail consists of a fin which is fixed to the fuselage or body and a rudder which is movable by the pilot. The fin (fixed portion) is a symmetrical airfoil which is in line with the center line of the fuselage. In steady flight, the rudder is



Principal parts of tail assembly.

stationary and approximately in line with the fin. The purpose of the vertical tail is to keep the air plane in line with the direction of motion or relative wind velocity.

In actual practice, the rudder may be slightly offset from the fuselage centerline to produce a small force on the fin to balance any slight difference in drag of the wings.

The span of the horizontal tail lies in a horizontal plane and usually consists of a stabilizer and elevator The stabilizer and elevator are symmetrical airfoil sections whose usual position is not in line with the fuselage centerline.

The principal purpose of the horizontal tail is to provide a balancing tail load so that the weight of the airplane, the lift, and the horizontal tail load are in equilibrium for a specific airspeed. *See* AIRPLANE

The elevator is moved by the motion of the pilot's control stick and can be moved rapidly, whereas the stabilizer can be moved slowly by means of a trim control which is independent of the control stick motion

Airplanes which operate at speeds greater than 80% of the speed of sound have tail surfaces which are sweptback 35–40°. *See* STABILIZER

Airplanes which operate at supersonic speeds usually have the horizontal tail sweptback and in one piece which is movable and is controlled by the motion of the pilot's control stick. Such a surface is often called a stabilator *See* AIRFRAME. [R G BO.]

Talc

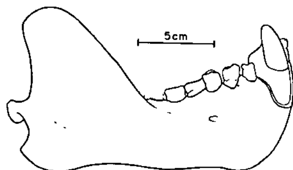
A hydrated magnesium silicate approximating in composition to $Mg_3Si_4O_{10}(OH)_2$ Chemical analyses frequently show percentages of calcium oxide CaO, aluminum oxide, Al_2O_3, and other oxides, but these arise mainly from impurities. Small amounts of Al_2O_3 may enter the talc structure. The structure is closely related to that of the micas, and consists of electricall neutral magnesium silicate

Taconite

The name given to the siliceous iron formation from which the high-grade iron ores of the Lake Superior district have been derived. It consists chiefly of fine-grained silica mixed with magnetite and hematite. As the richer iron ores approach exhaustion in the United States, taconite becomes more important as a source of iron. To recover the ore mineral in a usable form for the production of iron, taconite must be finely ground, and the magnetite or hematite concentrated by a magnetic or other process. Finally, the concentrate must be agglomerated into chunks of size and strength suitable for the blast furnace. *See* Iron (extraction from ore); Ore and mineral deposits; Ore dressing. [C.S HU.]

Taeniodonta

These extinct quadrupedal land mammals are known from early Tertiary deposits in the United States. One group, the Stylinodontinae, are adaptively convergent with many of the ground sloths, for in their evolutionary succession is seen the development of large size, of deep, massive lower jaws, of a reduced number of simplified, peglike upper and lower teeth with enamel remaining only as lateral bands, of curved and somewhat shearing



Side view of lower jaw of a middle Paleocene taeniodont, *Psittacotherium*. (After Matthew, 1937)

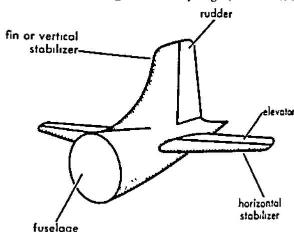
canines resembling rodent incisors, and of large, laterally-flattened claws. This succession leads from *Wortmania* of the early Paleocene to *Stylinodon* of the upper Eocene. The animals of the other group, Conoryctinae, ranging from *Onychodectes* of the early Paleocene to *Conoryctes* of the middle Paleocene, remain small to medium-sized and have less specialization in dentition and skeleton. Conoryctines developed enlarged canines, but the lower jaws were unspecialized and the cheek teeth, though simplified, remained low-crowned, enamel-enclosed, and cuspidate.

Taeniodonts probably arose from a generalized Cretaceous stock that would presently be assigned to the Insectivora. *See* Insectivora fossils [D.E S.]

Tail assembly

The assembly of the vertical tail, the horizontal tail, and a small section of the rear of the fuselage, as illustrated.

The vertical tail consists of a fin which is fixed to the fuselage or body and a rudder which is movable by the pilot. The fin (fixed portion) is a symmetrical airfoil which is in line with the center line of the fuselage. In steady flight, the rudder is



Principal parts of tail assembly.

stationary and approximately in line with the fin The purpose of the vertical tail is to keep the airplane in line with the direction of motion or relative wind velocity.

In actual practice, the rudder may be slightly offset from the fuselage centerline to produce a small force on the fin to balance any slight difference in drag of the wings

The span of the horizontal tail lies in a horizontal plane and usually consists of a stabilizer and elevator The stabilizer and elevator are symmetrical airfoil sections whose usual position is not in line with the fuselage centerline.

The principal purpose of the horizontal tail is to provide a balancing tail load so that the weight of the airplane, the lift, and the horizontal tail load are in equilibrium for a specific airspeed. *See* Airplane.

The elevator is moved by the motion of the pilot's control stick and can be moved rapidly, whereas the stabilizer can be moved slowly by means of a trim control which is independent of the control stick motion.

Airplanes which operate at speeds greater than 80% of the speed of sound have tail surfaces which are sweptback 35–40° *See* Stabilizer.

Airplanes which operate at supersonic speeds usually have the horizontal tail sweptback and in one piece which is movable and is controlled by the motion of the pilot's control stick Such a surface is often called a stabilator *See* Airframe. [R G.BO.]

Talc

A hydrated magnesium silicate approximating in composition to $Mg_3Si_4O_{10}(OH)_2$ Chemical analyses frequently show percentages of calcium oxide, CaO, aluminum oxide, Al_2O_3, and other oxides, but these arise mainly from impurities. Small amounts of Al_2O_3 may enter the talc structure The structure is closely related to that of the micas, and consists of electri

rs bonded together by weak, secondary valen- The mineral is therefore extremely soft (hard- 1 on Mohs scale) and possesses a perfect ... layers are

E MINERALS; SOLID-STATE CHEMISTRY.

he word talc or soapstone is applied also to sive materials which may contain many other eral constituents than talc itself. The more com-) associated minerals are other magnesian sili- s such as serpentine, chlorite, tremolite, and carbonate minerals, calcite, dolomite, and mag- ite. Steatite is a term applied to relatively pure pact and massive materials containing mainly mineral talc; steatites should contain not more n about 1.5% CaO, 1.5% combined FeO, Fe_2O_3, 1% Al_2O_3. *See* SOAPSTONE.

Talc occurs as a secondary mineral resulting m the hydration of magnesium bearing rocks d the alteration of minerals such as pyroxenes, phiboles, and olivine. Fibrous talc is often sely associated with, and probably derived from, molite, and the fibrous character is inherited om the parent mineral. The major talc-producing ates are New York, California and North Caro- a. The major talc-producing areas of the world e the United States, France, Italy, Austria, and pan (precise data for the U.S.S.R. and China e not available).

Talc is a widely used raw material. Figures for lc sold or used by the producers in the United tates in 1957 show that over 30% went into ce- amic applications, 20% was used in paints, and bout 10% for insecticides. The remainder was sed in a wide variety of applications. In the ce- amic industry, talc is used in many whiteware odies, tableware, electrical porcelain, high fre-

may also prevent crazing. In paints, talc is used as an extender and as a pigment. It is used as a filler for paper, rubber, and asphalt. In the cosmetics in dustry, it is used in toilet powders, soaps, and creams. Massive talc is cut into slabs and used for laboratory tables, sinks, sanitary appliances, acid tanks, electrical switchboards, mantels, and hearth stones.

[G.W.BR.]

Tall oil

A by product from the sulfate process for making cellulose. Tall oil and its derived products find in dustrial applications in the production of adhe- sives, binders, wetting agents, soaps, driers, emulsi- fiers, flotation agents, linoleum, printing inks, tex- tile finishes, and varnishes.

Pine wood is digested under pressure with alkali and sodium sulfate. Rosin acids and fatty acids are extracted from the pulp as soaps. So-called sulfate turpentine is also recovered during the treatment. Addition of mineral acids to the soaps of the rosin acids and fatty acids, and subsequent refin- ing gives a mixture of rosin and fatty acids known as liquid rosin, or tall oil. One ton of pulp yields as much as 100 lb of tall oil. Over 350,000 tons of crude tall oil was produced in 1958 in the United States. Tall oil yields rosin and fatty acids at lower cost than other sources.

Refining of tall oil by fractionation began on an extensive scale in 1942. Details vary from plant to plant, but all basically use injected steam and re- duced pressures to keep rosin and fatty acids from decomposing. The crude rosin and fatty acids are further up-graded to fulfill specific requirements.

About 90% of tall oil is composed of acidic ma- terials, containing approximately equal parts of rosin acids and fatty acids. The remainder is a com- plex mixture of fatty acid esters, sterols, higher alcohols, hydrocarbons, and decomposition prod- ucts resulting from the prolonged high tempera- tures used to distill the oil. The fatty acids are mainly oleic and linoleic acids. Smaller amounts of linolenic and palmitic acids are also present.

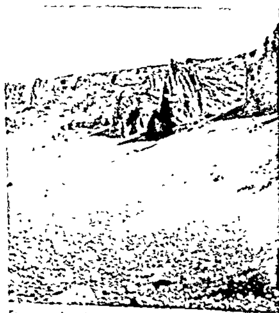
Modern methods of refining have produced grades of rosin and fatty acids each containing as little as 3% of the other. Treatments have been de- veloped to remove odor and color. Purification of tall oil by partition between immiscible solvents has been explored. If tall oil is cooled, rosin acids will crystallize. Recrystallization from suitable sol- vents such as methanol gives pure rosin acids.

Attempts are being made to develop tall oil rosin as a possible substitute for gum and wood rosin. Future research will determine the feasibility of this replacement. *See* DRYING OIL; ROSIN; TURPEN- TINE, WOOD CHEMICALS.

[F.L.S.]

Talus

A heap or sloping sheet of loose rock waste at the base of a cliff or steep slope. Talus and its English equivalent, scree, are terms properly applied to the



Fragmental rocks on north slope of Mud Creek Canyon, above the timber lines. Siskiyou County, California (USGS)

entire form and not to the fragmental material itself, which is rock waste or sliderock. Where the supply of rock waste is funneled downward through a notch or channel, a conelike talus usually develops.

Chemical weathering of susceptible layers or zones in a cliff weakens the support of overlying rock. Recurrent temperature changes and the freezing of water in narrow cracks tend to force joint blocks outward from their original positions. Loosened by these processes, blocks topple from the cliff and fall to the talus, where they slide or bound to a position of rest.

The upper part of a talus is characteristically steep, consists of coarse, angular rock waste, and is easily set in motion by falling blocks or by climbers. In cold climates its interstices may contain snow or ice for much of the year. The lower part is usually less steep and may be partly filled with finer rock debris and soil on which vegetation can gain a foothold. [C.F.S.S.]

Tanager

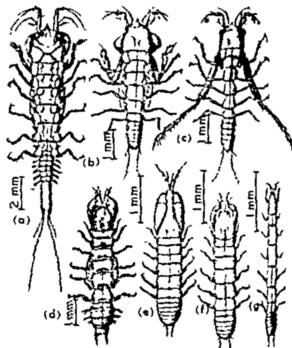
Any of many species of moderate-sized, blunt-billed perching birds of the family Thraupidae, all in the Western Hemisphere. Tanagers are often brightly colored, including brilliant red, orange, yellow, and blue species. The females are usually relatively dull in color. Most of the tanagers occur in tropical America, especially Central America. Four tanagers, all in the genus *Piranga*, reach the limits of the United States. The scarlet tanager, *P. olivacea*, is found over eastern Canada and the eastern United States; the male is a brilliant scarlet, with black wings and tail. The summer tanager, *P. rubra*, found from Delaware to Nevada and southward, is a uniform, dull red. The western tanager, *P. ludoviciana*, is yellow with a black tail, black wings and a red face; it is found throughout the western United States. The hepatic tanager, *P. flava*, of the Southwestern mountains, is similar to the summer tanager, but darker. *See* PASSERIFORMES. [J.D.B.]



The scarlet tanager, *Piranga olivacea*; length to 7½ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Tanaidacea

An order of the eumalacostracans of the superorder Peracarida, derived from the genus *Tanais* Audouin and Milne-Edwards 1829. These animals have a world-wide distribution and with few exceptions are marine. They occur from the shore down to abyssal depths. They are free-living and benthonic. The order is divided into 2 suborders with 5 families, 44 genera, and about 350 species. The body is linear, more or less cylindrical or dorsoventrally depressed (Fig. 1). Thoracic segments 1 and 2 are fused with the head, forming a carapace enclosing a respiratory chamber on each side, and the last abdominal segment is fused with the telson. The left mandible has a lacinia mobilis, while this structure may be present or absent on the right mandible. Eight pairs of thoracic legs are present, of which the first pair are maxillipeds, the second pair chelipeds, and the following six pairs pereiopods. The third pair in *Monokonophora*, as a rule, is fossorial, and the chelipeds and first pair of pereiopods usu-



Tanaidacea (After G. O. Sars). (a) *Apseudes spinosus* (M. Sars), female; (b) *Sphyropus anomalus* G. O. Sars, female, (c) *Sphyropus anomalus* G. O. Sars, male; (d) *Tanais cavolini* Milne-Edwards, female, (e) *Heterotanais oerstedi* (Kroyer), female; (f) *Heterotanais oerstedi* (Kroyer), male; (g) *Leptognathia filiformis* (Lilljeb.), female.

ally have vestigial exopodites. Pleopods may be present or absent, and the uropods are filiform.

The nervous system consists of a brain, a subesophageal mass, and a ventral chain. Eye lobes are present and sessile, with or without visual elements. Antennulae, especially in the males, have aesthetes. Sense hairs or sense spines are found on the segments and legs.

In the reproductive organs, the gonads are double. The oviducts open laterally at the base of the fourth pair of pereiopods. The vasa deferentia have a common vesicula seminalis, which is ventromedian on the last thoracic segment. Hermaphroditism, protandry, and protogyny can occur, and sex dimorphism is common. The antennulae are always different in the two sexes. Differences can occur also in the shape of the head, the mouthparts, the chelipeds, the first pair of pereiopods, and the [illegible] glands.

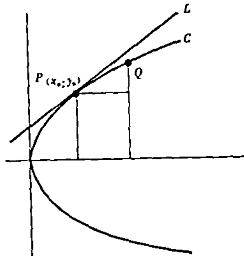
The females produce several broods, each preceded by a molting, by which the exterior morphology may undergo considerable alterations. The eggs develop in an incubatory pouch formed by one or four pairs of oostegites in which the embryos are dorsally flexed. The newly hatched larva lacks the last pair of pereiopods and the pleopods. The young probably undergo four larval (the manca) stages. See PERACARIDA; SEXUAL DIMORPHISM.

[K. L.]

Bibliography: K. Lang, The postmarsupial development of the Tanaidacea, *Arkiv Zool.*, 4:409–422, 1953; K. Lang, Neotanaidae nov. fam., with some remarks on the phylogeny of the Tanaidacea, *Arkiv Zool.*, 9:469–475, 1956; K. Lang, Protogynie bei zwei Tanaidaceen-Arten, *Arkiv Zool.*, 11:535–540, 1958; R. Siewing, Besteht eine engere Verwandschaft zwischen Isopoden und Amphipoden? *Zool. Anz.*, 147:166–180, 1951; R. Siewing, Morphologische Untersuchungen an Tanaidaceen und Lophogastriden, *Z. wiss. Zool.*, 157:333–426, 1954; T. Wolff, Crustacea Tanaidacea from depths exceeding 6000 meters, *Galathea Repts.*, 2:187–241, 1956.

Tangent

A term describing a relationship of two figures (usually of the same dimension) in the neighborhood of a common point. The figures are tangent at a point P if they touch at P but do not intersect in a sufficiently small neighborhood of P. To be more precise, if P denotes a point of a curve C, a line L is a tangent to C at P provided L is the limit of lines joining P to a variable point Q of C, as Q approaches P along C (that is, for Q sufficiently close to P, the line PQ is arbitrarily close to L). If curve C has equation $y = f(x)$ and point P on C has coordinates (x_0, y_0), it is shown in the



Line L tangent to curve C at P.

calculus that the slope of the line tangent to C at P is the value of the derivative $f'(x)$ for $x = x_0$. See CALCULUS, DIFFERENTIAL AND INTEGRAL.

[L. M. BL.]

Tangerine

A name applied to certain varieties of a variable group of loose-skinned citrus fruits belonging to the species *Citrus reticulata*. Although mandarin and tangerine are often used interchangeably to designate the whole group, tangerine is applied more strictly to those varieties having deep orange or scarlet rinds whereas the term mandarin is more properly used to include all members of this quite variable group of citrus fruits. See MANDARIN.

In the United States the main tangerine varieties are Dancy, grown widely in Florida on 23 000 acres, and Clementine, also called Algerian tangerine, grown chiefly in California and Texas. A number of hybrids, both spontaneous and man-made, are delicious and popular fruits. Important among these, and planted quite extensively, are the tangelos, which are hybrids between Dancy tangerine and grapefruit.

Dancy and Clementine tangerine trees are medium-sized with rounded tops bearing leaves pointed at the tip and rounded at the base, and having narrowly winged petioles. The flowers are white and small. The fruits are deep orange, loose-skinned, medium-sized to small, and possess small seeds with green cotyledons. Tangerines are easily peeled and eaten as fresh fruit. The average annual value of the tangerine crop in the United States for the past 10 years approximates $8,500,000. See FRUIT (TREE); see also FLOWER (BOTANY); FRUIT (BOTANY); LEAF (BOTANY); SEED (BOTANY).

[F. E. G.]

Tank

A container for storage of liquids or gases. A tank may be constructed of ferrous or nonferrous metals or alloys, reinforced concrete, or wood depending upon the use for which it is to be built. Tanks rest-

ing on the ground have flat bottoms; those supported on towers have either flat or curved bottoms. Standpipes, which are usually cylindrical shells of either steel or reinforced concrete resting on the ground, are frequently of great height and comparatively small diameter. They are built to contain water for a distribution system, and height is required to maintain pressure in the system. Tanks for other liquids and for gases, where storage is more important than pressure, are generally lower and of greater diameter.

Steel is usually the preferred material for standpipes because of the difficulty of securing watertightness in concrete shells with relatively high head. Unless painted regularly, steel standpipes are subject to rapid deterioration. Well-constructed concrete tanks need no surface treatment and are easily maintained.

Distribution-system pressure requirements limit the allowable fluctuation in water level in a standpipe or elevated tank to 25–30 ft. Therefore, unless a standpipe is located on high ground, only that volume of storage above the elevation required to give the necessary pressure is available for use. Elevated tanks become economical when the cost of the tower is less than the cost of the supporting portion of the standpipe below its useful head.

Elevated tanks for water storage are built of either steel or aluminum with storage capacities up to 2,000,000 gal. The higher initial cost of an all-aluminum tank may be offset by the need for less maintenance.

Elevated tanks with diameters less than about 50 ft usually have hemispherical bottoms. Bottoms of ellipsoidal or radial-cone shape are used on tanks of larger diameter. The roof may be conical or dome-shaped.

Prestressed-concrete tanks in which the concrete is in compression even when the tank is filled with water are constructed by a process used extensively in the United States. Wire is wrapped around a cylindrical concrete tank wall by a wire-winding machine suspended from a track placed on top of the wall. The wires are stressed to approximately 140,000 psi and are finally covered with a thin coating of pneumatically applied mortar. Prestressed-concrete tanks are watertight and require very little maintenance.

Molded plastic tanks in cylindrical and rectangular shapes and plastic liners for metal tanks are widely used in the chemical industry. [C.N.G.]

Tank circuit

The term tank circuit refers to an inductor and capacitor in parallel. The term is quite often used to denote the parallel resonant circuit in the output stage of a radio transmitter, but it has been applied to any parallel resonant circuit. In many cases the inductance in the tank circuit is one winding of a two-winding, air-core transformer. The secondary is connected to some load, such as an antenna. Power is delivered from the source to the load through the tank circuit, with an effort usually being made to adjust the parameters for maximum power transfer. *See* IMPEDANCE MATCHING; RESONANCE (ALTERNATING-CURRENT CIRCUITS); TRANSFORMER.

Since the tank circuit is a parallel resonant circuit, the parameters can be chosen so that at a desired frequency the voltage across the tank circuit will be a maximum. In radio transmitters this would be done by varying the capacitance, but in other situations, such as oscillators, the inductance could be varied by means of a tuning slug in the coil. *See* OSCILLATOR.

Tank circuits have an important role as a plate load in Class C amplifiers, where the plate current flows for only a small fraction of a cycle. If the damping in the circuit is small and the circuit is being excited at its resonant frequency, then the plate-current impulses will produce a sustained sinusoidal voltage across the tank. If a load is transformer-coupled to the tank, the voltage across the load will be sinusoidal. *See* AMPLIFIER. [H.F.K.]

Tannin

A generic term for a widely occurring group of substances of vegetable origin, capable of rendering raw hides into leather. They are obtained from various plant sources; common tannin (tannic acid) occurs in oak gallnuts (Turkish nutgall contains 50–60%, Chinese nutgall about 70%); tannins are also present in tea, sumac, oak bark (the word tan itself means oak bark), and mangrove bark. Tannin from the latter source is known as cutch, and is produced on a large scale, especially in Malaya.

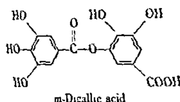
The usual method of preparation involves breaking or crushing of the bark or gallnuts into small pieces; these are then washed and boiled with water until the tannin has been extracted. After separation of insoluble matter, the thick, reddish-brown, viscous extract is evaporated, leaving the crude tannin as a hard cake. Purification may be effected by extracting the crude material with an alcohol-ether mixture; evaporation deposits the tannic acid as a colorless, noncrystalline mass. Tannic acid may also be prepared by heating gallic acid with phosphorus oxychloride.

Substances capable of tanning, and hence called tannins, are often of greatly different chemical structure; all tannins, however, have the property of converting the gelatin of hides into insoluble nonputrefying material, thus changing the hide into leather. In general, tannins are noncrystalline when solid, but readily soluble in water or alcohol to give colloidal solutions that are strongly astringent and therefore useful in medicine. Tannins have long been used in compounding inks, because they form greenish-black or bluish-black colors with ferric salts.

Tannins may be divided into three main classes: (1) condensed tannins that cannot be hydrolyzed either by acids or enzymes (these include the acacatechin and isoacacatechin tannins and the gam-

catechin tannins, all contain highly substituted [ph]loroglucinol nuclei); (2) hydrolyzable tannins, for example, gallotannins, ellagitannins, and caf[fet]annins; and (3) tannins of unclassified nature.

Gallotannin, from which is obtained the tannic [aci]d of commerce and medicine, is present in oak [ga]lls. It is a mixture of the gallic acid esters of [gl]ucose, one of which is pentadigalloylglucose, the gallic acid moiety having the following struc[tu]re:



m-Digallic acid

These esters are called depsides.

Tannic acid, USP, is a mixture of compounds of [g]allotannin type It is a light yellow powder of very astringent taste, much used in styptic preparations and ointments The aqueous solution of tannic acid is used in treating burns, because it precipitates the burned protein, forming a nonputrefying, protective layer under which new tissue can grow. *See* LEATHER AND FUR PROCESSING

[E.B.R.]

Tantalite

A mineral with composition $(Fe,Mn)Ta_2O_6$, an oxide of iron, magnesium, and tantalum Columbium (niobium) substitutes for tantalum in all proportions, a complete series extends to columbite $(Fe,Mn)Cb_2O_6$ Pure ...

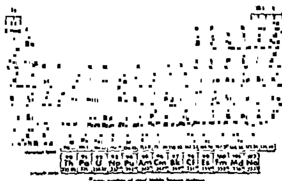
... is common in short prismatic crystals. There is perfect side pinacoid cleavage The hardness is 6 the specific gravity 7.95 (pure tantalite) The luster is submetallic and the color iron-black Tantalite is the principal ore of tantalum It is found chiefly in granite pegmatites and as a detrital mineral, in some places in important amounts having weathered from such rocks The chief producing countries are the Belgian Congo and Nigeria *See* NIOBIUM, TANTALUM

[C.S.HU.]

Tantalum

Chemical element number 73, tantalum, Ta, has atomic weight 180.95 Tantalum is a member of the fifth subgroup of the periodic table and is in the 5d transitional series. Its valence-electron configuration is $5d^36s^2$, which results in a maximum oxidation state of 5+ Oxidation states of II, III, and IV also are reported The Goldschmidt radius for the Ta^{5+} ion is 0.73 A Tantalum is found in nature with its lower-atomic-numbered homolog niobium in the following ores: columbite or tantalite, $(Fe,Mn)(Nb,Ta)_2O_6$, depending upon which element is predominant; euxenite, $Y,Ca,Ce,U,Th,(Nb,Ta,Ti)_2O_6$, and by itself in microlite, $(Na,Ca)_2Ta_2O_6(O,OH,F)$ Tantalum occurs in the earth's crust to the extent of 2.1×10^{-4}%. In 1957, the principal source of niobium and tantalum concentrates was Africa. The United States supplied only a small percentage of the world production.

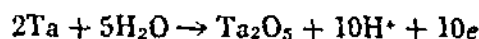
More than 50% of the tantalum metal produced in 1957 was used in the manufacture of capacitors. Tantalum also has been studied as a container for the uranium-bismuth slurry in the liquid-metal-fuel reactor experiment and in the Los Alamos molten-plutonium reactor experiment. It is also used for heat-transfer surfaces in chemical production equipment, especially where extraordinarily corrosive conditions exist. Because of its high cost, however, it has had limited application. Its chemical inertness has led to dental and surgical applications



Metallurgical extraction. The coexistence of niobium and tantalum in ores and the similarity in their properties requires a fractional method of separating one from the other. Up to recently, the fractionation had been accomplished by crystallization of the salts from fluoride systems, taking advantage of the fact that niobium tends to form oxygenated species such as $NbOF_5^{--}$ under concentration conditions at which tantalum crystallizes as the salt of the TaF_7^{--} ion The niobium salt is more soluble than the tantalum salt Recently developed methods of separation involve liquid-liquid extraction procedures In one such process, the aqueous solution of metal ions containing 0.5 N HCl and 3.3 N HF is treated with an organic solvent such as methyl isobutyl ketone The tantalum species are distributed preferentially into the organic phase, with the niobium species concentrating in the aqueous phase The organic phase is back extracted with water to recover the tantalum. Tantalum and niobium are recovered as the oxides from the aqueous phases by complexing the fluoride with boric acid, and precipitating the oxide with aqueous ammonia The precipitate obtained from the initial aqueous layer is 98% niobium oxide, and that from the organic layer is 99.5% tantalum oxide In the absence of hydrofluoric acid and at very high hydrochloric acid concentrations, the niobium concentrates in the organic phase and the tantalum remains largely in the aqueous phase. Other methods based on anion exchange have also been reported *See* SOLVENT EXTRACTION.

Properties. Metallic tantalum can be prepared by the electrolysis of fused K_2TaF_7 or by the re-

duction of the oxide with active metals or carbon. A powder is obtained which, after being washed, is compacted into bars in presses and then sintered in a vacuum furnace with the bar acting as the heating element. The bar is then cold rolled to sheets or wire. Tantalum metal has a density of 16.6 g/cm³ and crystallizes in the body-centered cubic system. It has a melting point of 2996°C and a boiling point greater than 4100°C. It has a cross section for capture of thermal neutrons of 21.3 barns. Although the standard oxidation potential for tantalum metal for the reaction



$$2Ta + 5H_2O \rightarrow Ta_2O_5 + 10H^+ + 10e$$

is 0.71 volts, the metal in actual practice is quite inert to acid attack except by hydrofluoric acid. It is very slowly oxidized in alkaline solutions. The halogens and oxygen react with it on heating, and at high temperatures, it absorbs hydrogen and combines with nitrogen and carbon to form carbides and nitrides.

The aqueous chemistry of tantalum is practically nonexistent except for the fluoride and very strong mineral acid solutions because of the insolubility of most of the tantalum compounds. The nature of the species in the strong acid solutions is not well understood. The reduced states of tantalum are not produced in an aqueous solution.

Principal compounds. The oxide, fluoride, chloride, bromide, and iodide of the V oxidation state are the most important binary compounds of tantalum. Tantalum(V) oxide has a melting point of 1470°C and is best made by heating the metal in oxygen. When the Ta^V oxide is dissolved in fused alkali hydroxides or carbonates, tantalates (TaO_4^{--}) result which are insoluble in water and hydrolyze to the oxide upon washing. Tantalates of the general composition $Ta_6O_{19}^{8-}$ which are soluble in water can be prepared. The hydrated oxide is precipitated from these solutions upon the addition of acid. The oxide is insoluble in all acids at moderate concentrations except in hydrofluoric acid; it is slightly soluble in concentrated mineral acids.

The pentahalides are all low-melting, low-boiling compounds which hydrolyze to give the oxide when placed in water. They are prepared by direct action of the halogen on the metal or by action of dry hydrogen halides on the metal.

Tantalum pentahalides, TaX_5

	Melting point, °C	Boiling point, °C
TaF_5	95.1	229.2
$TaCl_5$	220.0	239.3
$TaBr_5$	280.0	348.8
TaI_5	543 ± 0.5	496 ± 2.0

The chemistry of the reduced states of tantalum has not been thoroughly explored. Tantalum tetrachloride and tetrabromide have been prepared by reduction of the pentahalides by hydrogen in an electric discharge tube. The tetrabromide is known to disproportionate to the III and V oxidation states at 300°C, and the III state to the II and V states at still higher temperatures.

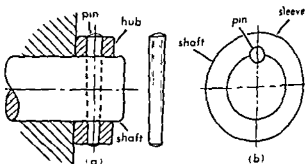
The important compound K_2TaF_7 may be formed by fusing tantalum pentafluoride with potassium hydrogen fluoride. It may be recrystallized from an aqueous solution containing hydrofluoric acid. It is thermally stable to temperatures beyond its melting point, ultimately decomposing to the original reaction products.

Analysis. Tantalum is determined in the presence of niobium by developing the peroxytantalate color in 96% sulfuric acid and measuring the absorbancy at 285 millimicrons (mμ); the niobium may be determined at 365 mμ. *See* NIOBIUM; TRANSITION ELEMENTS. [E.M.L.]

Bibliography: J. Kleinberg (ed.), *Treatise on Inorganic Chemistry*, vol. 2, 1956; N. V. Sidgwick, *The Chemical Elements and Their Compounds*, vol. 1, 1950.

Taper pin

Standard taper pins have a diametral taper ¼ in in 12 in. and are driven in holes drilled and reamed to fit. The pins are self holding and made of soft steel or are cyanide hardened. They are sometimes used to connect a hub or collar to a shaft (as illustrated). Taper pins are frequently used to maintain the location of one surface with respect to another.



Taper Pins (a) Connecting hub or collar to shaft (b) Connecting sleeve to shaft.

A disadvantage of the taper pin is that the holes must be drilled and reamed after assembly of the connected parts; hence they are not interchangeable. *See* COTTER PIN. [P.H.B.]

Bibliography. *Society of Automotive Engineers' Handbook*, revised annually.

Tapeworm

Any of over 1500 species belonging to the class Cestoda, phylum Platyhelminthes. Tapeworms are all parasitic in the intestines of vertebrates. Several are important parasites of man.

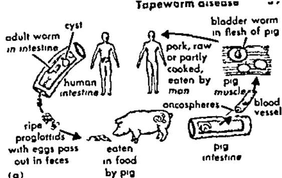
Structure. The adult tapeworm consists of a head, or scolex, which terminates in a number of segments, called proglottids. The rounded scolex, armed with suction cups and hooks, is the organ by which the tapeworm attaches itself to the intestinal lining of the host. The flattened proglottids are produced directly back of the scolex and, as soon as one has formed, it is pushed back by a new proglottid. Thus when the proglottid is mature, it is at

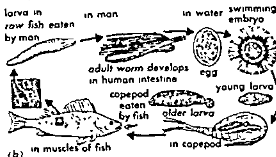
ie posterior end of the worm Each segment contains ovaries and testes. Tapeworms may be self-fertilizing within a proglottid, or cross-fertilization may occur between proglottids or between worms. When fully mature, each proglottid has a uterus crowded with thousands of eggs. In each segment there are also a pair of excretory canals, many flame cells, and three pairs of nerve cords. Virtually all these organs degenerate as the uterus becomes filled with eggs. Because they lack an intestine, tapeworms are unique among animals of this degree of organization Food is absorbed directly from the environment. There is no respiratory system, and respiration is accomplished in all the tissues by cellular gaseous exchange. Respiration is primarily anaerobic, but oxygen is utilized when available

Reproduction. While the life cycle of the beef tapeworm, *Taenia saginata*, is not typical of the cestodes, this tapeworm is such a common parasite that its life cycle is discussed below. The adult lives in human intestine When gravid, each proglottid drops off and leaves with the feces. The eggs are scattered with the disintegration of the proglottid. Development of the ovum, already underway, continues through the formation of a six-hooked embryo, the oncosphere, and then it enters a resting stage. If the egg at this stage is eaten by a cow or any other ungulate, the shell is digested off and the larva burrows into the veins or lymph vessels of the intestine where it is carried to muscle tissue In the muscle it encysts as the bladder worm, or cysticercus. Here it may remain dormant for months or years. If the raw or partially cooked meat is eaten by man, the outer wall of the cyst is digested away and the scolex of the young adult emerges. This attaches to the intestine wall and the development of proglottids begins The beef tapeworm attains a length over 30 ft and produces 2000 or more proglottids

The somewhat smaller pork tapeworm, *Taenia solium*, has a life history similar to that of *T. saginata*. The common larval host is the pig, but it may occur in other animals including dogs, cats, a . .

Another tapeworm harmful to man is the fish tapeworm, *Dibothriocephalus latus* This large species may become 60 ft long and have over 4000 proglottids It is especially dangerous, sometimes fatal, because eating a single undercooked fish may result in ingesting a large number of cysticerci An Old World species, it is now increasing in parts of the northern United States and Canada. In recent years, several deaths have been reported in North America as a result of infestations with this worm. Almost any fish-eating mammal may be infested with the adults Larvae develop first in copepod crustaceans and then complete development in a variety of fishes, including the northern pike walleye, yellow perch and trout.





(a) Life cycle of tapeworms. Pork tapeworm, *Taenia solium* (adapted from Buchsbaum). (b) Fish tapeworm, *Dibothriocephalus latus* (adapted from Kükenthal). (From T I. Storer and R L. Usinger, General Zoology, 3d ed, McGraw-Hill, New York, 1957)

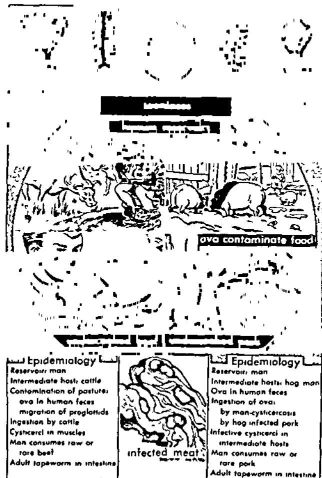
The hydatid worm, *Echinococcus granulosus*, attains its adult stage in dogs and related species. The larvae occur in many animals including man. This worm is serious because of its formation of large hydatid cysts, sometimes 2–3 in long, which cause pressure on vital organs, including the brain, and result in great pain or death. Although most common in Iceland and Greenland, it is cosmopolitan in distribution. About 500 cases have been reported in the United States. See CESTODA [J.D B.]

Tapeworm disease

The presence in the body of man of either the larval or adult stages of any of the following cestodes: *Hymenolepis nana*, *Taenia saginata*, *T. solium*, *Dibothriocephalus latus*, *Echinococcus granulosus*. See CESTOIDEA

Hymenolepiasis. A cosmopolitan parasite, *H. nana*, also occurs in rodents. Prevalence in man is usually under 1%, children being more commonly infected The cycle is direct, man to man or rodent to man Ingestion of the egg results in development of larvae, called cysticercoids, in cysts in the duodenal villi After the larvae burst out of the cyst, they attach to the small intestine and become adults The entire cycle takes a month. Heavy infection may result in nervous disorders Treatment consists of hexylresorcinol, acranil, or chloroquine given orally

Beef tapeworm. *T. saginata*, the beef tapeworm is acquired by ingesting the larval form of the parasite *Cysticercus bovis* that grows in the intramuscular connective tissue of herbivores. The adult, attached by the scolex or head to the small



Epidemiology of the taeniases. (*From T T Mackie, G W. Hunter, and C. B Worth, A Manual of Tropical Medicine, 2d ed , Saunders, 1954*)

to pass actively through the anus Diagnosis is made by recognizing these segments with their peculiarly branching uterus. When the extruded segment disintegrates, thousands of hardy, embryonated eggs are set free If eventually ingested by cattle, the cysticercus will develop and become infective after 2 months

No symptoms may be exhibited by parasitized individuals, although nervous disorders may arise. Emotional distress is the most serious complaint

Treatment consists of 0.6–0.8 g of atabrine orally preceded and followed by saline purging

Pork tapeworm. The epidemiology of *T solium*, the pork tapeworm, is essentially the same as that described for *T. saginata*, except that the parasite uses the hog as intermediate host The larva of *T. solium* is called *C. cellulosae*.

C. cellulosae may develop in man with serious results. Human cysticercosis is acquired by ingesting *T. solium* eggs from food or contaminated hands, or, in people harboring the adult worm, by regurgitation of segments into the stomach. The larva grows in connective tissue but exhibits an affinity for the central nervous system. There it may cause epileptiform symptoms or a syndrome suggestive of brain tumor, or other derangements. The prognosis is poor.

Fish tapeworm. *D. latus*, the fish tapeworm, inhabits the small intestine of man. While the Ba[ltic] and Scandinavian countries are the classical foc[i], other cold regions are also important ones. In f[act], wherever fresh-water fish are eaten raw, the dise[ase] is prevalent. Many mammals act as reservoirs.

The eggs of the worm embryonate and hatch [in] water. The ciliated embryo, called a coracidium, [is] ingested by water fleas (*Cyclops, Diaptomu[s]*) where it becomes a procercoid Many species [of] fish ingest the crustacean. In the flesh the la[rva] turns into a plerocercoid; in other hosts, into [a] sparganum. These in turn mature in the mamm[al].

Heavy infection may cause intestinal obstr[uc]tion and possibly toxemia. In less than 1% of [the] cases patients develop pernicious anemia, wh[ich] abates after successful therapy. Treatment is [the] same as for other large cestodes.

Echinococcosis. This is also known as hyda[tid] disease and is an infection by the larval form [of] *E. granulosus* in man or other intermediate ho[sts]. Adult *E. granulosus* are minute parasites of [the] small intestine of carnivores. Shepherd dogs [are] important epidemiologically. Sheep-growing cou[n]tries are most affected. The carnivore passes eg[gs] in feces so that grazing sheep and other herbivo[res] become parasitized with the hydatid larva. The [cy]cle is completed by carnivores preying on the[m].

Man becomes exposed by petting dogs or ing[es]tion of contaminated grass In 60% of cases [the] larva settles in the liver, but bone, lungs, and bra[in] may be affected. The type of cyst varies with loc[al]ization It is essentially a bladder full of liquid. [In] a year it is approximately 1 in in diameter, a[nd] may grow to a larger size. Cysts may be sterile [or] fertile If fertile, the cyst is bordered inside by [a] proliferative layer that gives rise to numberle[ss] scolices These are known as hydatid sand Secon[d]ary cysts may form within the parent one. T[he] symptoms are referable to a slowly growing tum[or]. Prognosis is good if surgical removal succeed[s], poor if secondary infection occurs and in inop[er]able cases. *See* PARASITOLOGY, MEDICAL [J.F.M.]

Tapir

Any of several species of primitive, odd-toe[d] hoofed mammals, all of the genus *Tapirus*, fami[ly]



The Malay Tapir, *Tapirus indicus*; length 8 ft. (*From P. Martin Duncan, ed., Cassell's Natural History, Cassell & Company, Ltd., London*)

Tapiridae, order Perissodactyla. Tapirs are usually found near water in the tropical forests of the Americas, the Malay peninsula, Sumatra, and Borneo. They are shy, nocturnal animals, at home in the water or on land, they feed upon the succulent vegetation of wet jungles.

Tapirs have the upper lip modified into a short, flexible proboscis; they are characterized by a thick, sparsely-haired skin, small ovate ears, and a short tail. They have four toes on the front feet and three on the hind feet. The flesh of the Malayan tapir is not considered desirable food by man, but the meat of the four American tapirs is of excellent quality. See PERISSODACTYLA. [J.D.B.]

Tardigrada

A class of microscopic, bilaterally symmetrical invertebrates which are generally less than 1 mm in length. About 350 species are known. Commonly called water bears, bear animalcules, or urslets, they are world-wide in distribution and are found in all habitats.

Systematic position of the Tardigrada (Marcus)

- Superphylum Articulata
  - Phylum Annelida
  - Phylum Oncopoda
    - Subphylum Pentastomida (Linguatulida)
    - Subphylum Malacopoda
      - Class Onychophora
      - Class Tardigrada
        - Order Heterotardigrada
          - Suborder Arthrotardigrada
          - Suborder Echiniscoidea
        - Order Eutardigrada
  - Phylum Arthropoda

An independent but associated origin of the two malacopodan classes from polychaete-like ancestors may be presumed.

Anatomy. The tardigrade body consists of an anterior prostomium and five segments. The mouth is located in the prostomium in a centroterminal position. A soft, nonchitinous cuticle surrounds the body and lines the fore- and hindgut. The cuticle may be smooth or sculptured and forms innervated cephalic appendages and spines on the trunk and legs. Four pairs of ventrolateral legs arise from the trunk and terminate in claws or other modified structures. A pair of oral glands and stylets are present which are partly cuticular and partly chitinous. Both the pharynx and esophagus are muscular structures. The digestive tract is tubular and more or less lobed due to the presence of diverticular dilations. In the eutardigrades, separate anal and genital openings occur, while in the heterotardigrades there is a single anogenital opening, the cloaca. The sexes are separate and the gonads are unpaired dorsal sacs with paired gonoducts in the male and in theory also in the female.

Histologically, the musculature of the body, legs, and mouth is of the smooth type.

The brain is a voluminous supraesophageal ganglion connected to a subesophageal ganglion by a pair of connectives. The brain has 2–3 inner and 2 outer lobules. Each of the latter, in most species except Arthrotardigrada, has one eye spot. This structure is an ocellus which is cupped-shaped, and consists of one retinal and one pigmented cell which may be black, red, or pigment free. The ventral nerve cord, with four ventral ganglia, is a continuation of the subesophageal ganglion. Food storage cells float in the spacious body cavity, the coelom, which lacks a parietal or visceral peritoneum in the adult. Circulatory and respiratory structures are lacking.

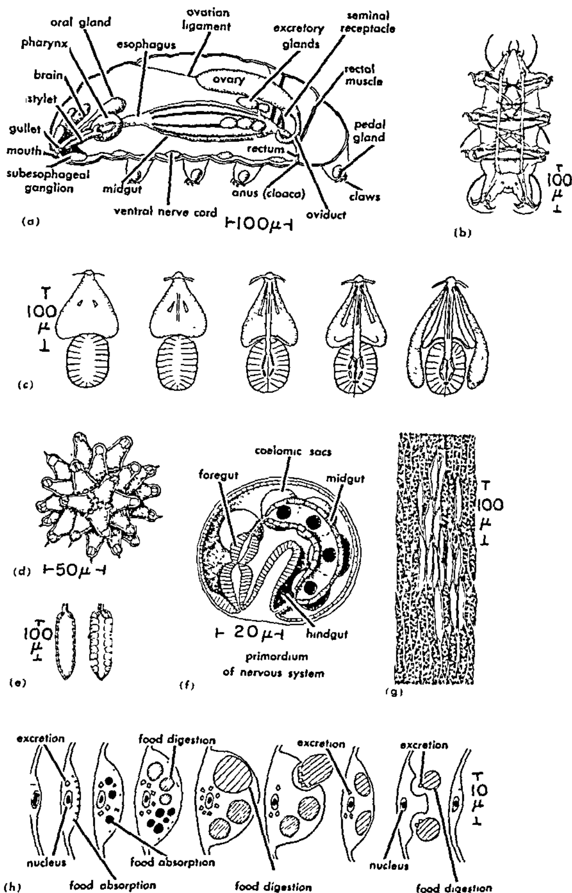
These animals exhibit the phenomenon known as cell constancy. The number of epidermal cells is the same in all species of a genus. The pharynx has 27 epithelial and 24 myoepithelial cells whose number is also constant in all genera except in *Milnesium* which has 24 and 39.

The lumen of the pharynx is triradial. Body muscles are metameric, and comprise dorsal, ventral, and lateral groups. Each muscle is either a single fibrillar, uninucleated cell or a chain of such structures. The number of myocytes is constant in all individuals of the same species. The muscles retract and bend the body, while during relaxation, pressure of the fluid in the body-cavity extends and stretches the animal. Storage cells of newly hatched *Echiniscoides* are constant in number and appressed to the epidermis; older cells multiply when dissociated from the epidermis.

Embryology. These animals lay eggs and development is direct. Embryonic development lasts 3–40 days, which varies according to the species and temperature. Fertilization, as known in the Eutardigrada, may be external while eggs are laid in the old cuticle during molting, or internal within the ovary. Internal fertilization occurs after ejaculation of sperm into the female cloaca. Some species may be parthenogenetic.

The oocytes mature with abortive oocytes serving as nurse-cells. The ovarian endothelium produces the shell (chorion), which is smooth or has processes (ornamentations), that are of taxonomic importance. The diploid number of chromosomes is 10–14 for species of Eutardigrada. Liberated eggs vary in diameter from 60 to 200 microns (μ). The number of eggs in a single oviposit varies from 3 to 30. Cleavage is total, nearly equal, and is irregular because of asynchronous division of the blastomeres.

Gastrulation, by coeloblastic delamination, occurs at the 80- to 100-cell stage. Primordial germ cells are recognizable in the primary entoderm. The enterocoelous origin of the entomesoderm results in the formation of 5 pairs of coelomic sacs. The 4 anterior pairs give rise to muscles and storage cells, while the posterior pair unite to form the gonad from which gonoducts develop. The nervous system [illegible] amination. [illegible] gut. [illegible] bout 51 μ. [illegible] is about 1200 μ. Very young animals swell as a result of water intake. Fewer claws and appendages are characteristic of the



Tardigrada. *(a) Macrobiotus*, principal organs; *(b)* muscles, ventral view *(from Johanna Müller, 1935)*; *(c)* renewal of buccal apparatus in 5 consecutive days, *(d)* egg with processes, *(e)* gut before and after absorption; *(f)* development of coelomic sacs; *(g)* 16 moss cells emptied by tardigrade; *(h)* digestive processes in gut cells.

nile. These animals ordinarily grow to 3–5 times r initial size and the immature *tardigrades* may eggs.

lolting. *During their* active life of 18 months, ligrades *molt* about 12 times. They are unable eed in the 5–10 days of molting. Molting begins n the animal passes to the simplex stage, which haracterized by expulsion of the buccal cuticu- parts. During molting, the epidermis secretes a v cuticle, and epidermal foot glands renew the s or claws. The oral glands shorten and then generate stylets and bearers. In *Diphascon*, *epi-* thum of the gullet, pharynx, and *rectum replace* ... certain conditions.

rdigrada,

new one

ickens to form a cyst. Hunger and perhaps un- ual warmth seem to elicit encystment. Histolysis thin a cyst *appears improbable*. It is possible at consumption of stored food causes the tardi- ade to leave its cyst upon formation of the third ticle. Molting is completed when the newly rmed stylets pierce the old cuticle to allow the nimal *to escape. Duration* of active *life*, without ncystment, *for as long as* 30 *months* has *been re-* orded for these animals.

Physiology. Tardigrada live as active forms, ...nded by a

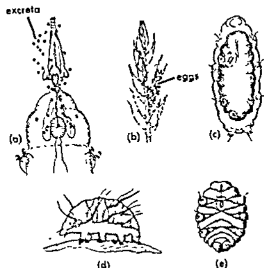
means of a pumping pharynx As a result, many ingested *chloroplasts* are fragmented The foregut

droplets containing plant matter. These cells contain digestive ferments. Fats, starch, and glycogen have been demonstrated in the storage cells

Sometimes tardigrades, especially *Milnesium*, pierce the cuticle of rotifers, nematodes, other tardigrades, *and* nauplii to ingest their soft parts Active animals can fast as long as 5 weeks In active

(part of *pigments*) during the formation of the new cuticle.

a loss of oxygen from the water The animal responds by contraction and loss of body water. Dried eggs also survive. Moistened animals usually revive, but anabiosis and revival *cannot* be repeated indefinitely. Experimentally, the *limit of survival* was 14



Tardigrada (a) Expulsion of buccal apparatus and excreta, (b) eggs *fastened to moss leaf; (c) cyst; (d)* scutechiniscid *walking; (e) barrel.*

months, it is found that a tardigrade might live for 67 years, hence the name *Macrobiotus*. The limit of life during anabiosis does not depend on the amount of stored food, since those with a reserve food supply do not differ in anabiotic capacity from those without one. These facts imply that the cause of death must be endogenous

A 100-year-old discussion on anabiotic animals was characterized by two opposing views One argument maintained that *anaerobiosis in these animals was actually a very slow metabolism or "vita minima"* The other extreme held the view that life stopped without death supervening, and the organism was compared to a wound clock which had stopped ticking A. Pigoń and B Węglarska settled the matter in 1953 by establishing a vita minima. They found that cysts were resistant to 92°C heat in air for 1 hour. The cysts withstood cold at temperatures to −192°C for 20 months, and to −272°C for 8½ hours. In liquid air or helium used *to study low-temperature effects, cysts are not completely dry, but contain water within and between* their cells

Tardigrades contract to the characteristic barrel-shape when dissolved oxygen decreases in the water They cannot maintain this state *in water*, however, and in 48 hours they become maximally distended and immobile. Such asphyctic tardigrades may continue to live for 5 days and can revive if oxygen and food are made available. Animals emerging from anabiosis pass *into* a state of asphyxy, since the cells have lost their osmotic *capacity* The arthrotardigrade, *Batillipes*, does not endure desiccation. For *Echiniscoides sigismundi*, about 42% of a population withstands dryness for 10 days, but none survived after 4 weeks. All were found to be alive in rain water after 3 days, and in *an asphyctic state.*

Movement in the *tardigrades is* varied. Scutechiniscidae walk with legs which *are nearly* perpen-

dicular, under the trunk, while *Tetrakentron* crawls. No species swims. *Echiniscoides* shows positive phototaxis, while the negative response of *Macrobiotus dispar* is a directed reaction. Eggs are frequently laid with positive thigmotaxis either between the leaflet and stalk of moss or in the shells of water fleas.

Ecology. Most species are widely distributed. Dissemination may be by wind, birds, and those terrestrial animals which transport tardigrade eggs and barrels. Other moss dwellers, such as rhizopods, rotifers, and nematodes are more easily transported than tardigrades. The tropics are especially poor in numbers of species, and no Scutechiniscidae have been found on the Antarctic continent

Echiniscoides sigismundi is the most euryokous species since it is world-wide in distribution and found both in littoral algae and at altitudes up to 1000 meters in the eastern Belgian Congo. In this environment tardigrade populations probably survive the dry periods Population density varies considerably In one sample of water containing *Enteromorpha*, 1 milliliter yielded 60 specimens.

A few scutechiniscids and several eutardigrades are limnetic organisms. The limnetic fauna is not sharply delimited from the terrestrial fauna, so that more limnetic than marine species are found, usually among algae, aquatic mosses, sand, and mud. Sandy beaches of soft-water lakes contain more specimens than do those of hard-water lakes. Some populations of permanently aquatic habitats cannot survive desiccation. Most species are eurythermal, having the ability to live through a wide range of temperature conditions. They are active between near-freezing temperature and 25–30°C One known species is continuously exposed to temperatures of 40°C

Most tardigrades are terrestrial. They are found among lichens, liverworts, densely growing soft-leaved mosses, and also in rather hard-leaved Pottiacea and Grimmiacea They are rarer in some ferns, lycopods, and phanerogams such as *Sedum*, *Saxifraga*, and *Haastia* The soil among fallen needles and leaves, will yield tardigrades. As many as 22,000 animals have been recovered from 1 gram of air-dried moss. An average population produces 354 kilograms of humus per hectare yearly, chiefly by means of fecal decomposition.

Fewer species are found in permanently wet or damp mosses than in land mosses. Oxygen supply is better in the land mosses and desiccation excludes many rival organisms

Tardigrades and their eggs are preyed upon chiefly by Amoebozoa, especially *Difflugia*. Nematodes attack the eutardigrades. Common pathogens of tardigrades are Phycomycetes (*Macrobiotophthora vimariensis*) and Microsporidia (*Pleistophora*). *See* CELL CONSTANCY; CLEAVAGE, EMBRYONIC; EUTARDIGRADA; HETEROTARDIGRADA; POLYCHAFTA. [E.M.]

Bibliography: E. Marcus, *Tardigrada*, in H. G. Bronn (ed.), *Klassen und Ordnungen des Tierreichs*, 1929; E. Marcus, *Tardigrada*, in F. Schulze, *Das Tierreich*, pt. 66, 1936; R. W. Pennak, *Fresh-water Invertebrates of the United States*, 1953; B. Petersen, The tardigrade fauna of Greenland, *Medd. Grønland*, 150, 1951; G. Ramazotti, I Tardigradi d'italia, *Mem. ist. ital. idrobiol*, 2:29–166, 1945; H. B. Ward and G. C. Whipple, *Fresh water Biology*, 2d ed., 1918.

Tarnished plant bug

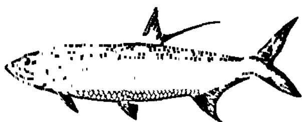
An insect, *Lygus lineolaris*, of the family Miridae, order Hemiptera. This insect is a pest which damages a wide variety of plants, including many of importance to man. It frequently damages the buds of fruit trees, small fruits, and flowers. It carries the bacterial disease fire blight of pears, and causes a condition called stopback in young peach trees by destroying the terminal bud of the young tree.

The tarnished plant bug is about 1/4 in. long, brownish, somewhat variegated with lighter shades, and with irregular black markings. It may produce as many as five generations a year. DDT and toxaphene are among the insecticides used in its control. There are several closely related species. *See* HEMIPTERA. [J.D.B.]

Tarpon

A large bony fish, *Tarpon atlanticus*. It is found in the Atlantic Ocean from Long Island south to Brazil, but occurs most commonly in the Gulf of Mexico and the Caribbean Sea.

The tarpon is highly prized by sport anglers. Although edible and sold for food in Latin America, it is considered primarily a sports fish. The tarpon attains a weight of 350 lb, however, specimens over 100 lb are rare. It responds with a spectacular fight when hooked.



The tarpon, *Tarpon atlanticus*, length to 8 ft. (*From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

The tarpon is readily recognized by its size, and the compact dorsal fin The last ray is elongated and situated in the middle of the back. It has a long, sickle-shaped anal fin; large and deeply forked caudal fin, and large, silvery scales. The scales are frequently sold as souvenirs at ports where tarpon fishing is promoted *See* CLUPEIFORMES [J.D.B.]

Tartaric acid

Tartaric acid, HOOCCHOHCHOHCOOH, exists in two enantiomorphic forms, one meso, and one racemic; *d*- and *l*-tartaric acids have melting point 167–170°C, and $\alpha_D^{20} = \pm 12°$; mesotartaric acid has melting point 140°C, and is optically inactive because of internal compensation; racemic tartaric

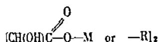
, melting point 206°C (decomposition), and
lly inactive See OPTICAL ACTIVITY.
ordinary form, *d*-tartaric acid is obtained
rmented grape juice as potassium hydrogen
(argol or cream of tartar). Racemization
taric acid with hot sodium hydroxide, or ni-
d oxidation of mannitol or mucic acid, gives
c tartaric acid Resolution of racemic tar-
cid via *Penicillium glaucum*, the cinchonine
icine salts, or *l*-bornyl hydrogen ester, fur-
l tartaric acid

rate) *See* CARBOXYLIC ACID; TARTRATE.

[E B R.]

rate

emical compound that is a salt or ester of tar-
acid, and has the general formula



$$[CH(OH)\overset{O}{\overset{\|}{C}}—O—M \quad \text{or} \quad —R]_2$$

e are formed by replacing the carboxylic hy-
gens of tartaric acid by a metal (salt) or or-
c radical (ester) One or both of the carbox-
hydrogens can be replaced to produce an ex-
sive series of salts, esters, and mixed (double)
s Tartrates exist in three isomeric forms, two
ng optically active and one inactive (*see* OPTI-
ACTIVITY). Salts are commonly made from crude
tars, obtained as a by-product of the wine-mak-
industry Many have practical applications, as
cream of tartar (monopotassium salt), Rochelle
ts (sodium potassium salt), tartar emetic (po-
ssium antimony salt), and medicines and textile
eing (calcium salt) Esters are not used exten-
ely. *See* ROCHELLE SALT; TARTARIC ACID

[E.H.H.]

aste

ne of the chemical senses Taste is not a simple
ensory pattern It depends not only on the specific
aste receptors of the tongue, but also upon other
ensations supplied by the chemical, tactile, warm
nd cold receptors in the mouth, and the smell re-
ceptors in the nose Taste and taste discrimination
are reduced if the sense of smell is suppressed, for
example, by a cold.

Anatomy. *In man* and mammals, taste buds are
located on the tongue in the fungiform, foliate, and
circumvallate papillae and in adjacent structures
of the throat. Taste buds are goblet-shaped clusters
of cells with a small pore opening to the mouth
cavity Two types of cells have been described in
the taste bud, the thicker supporting cells, and the
more slender, *gustatory cells* from which a fine ter-
minal hair projects into the taste pore, although
some workers question this dichotomy Individual
sensory nerve fibers entwine about one or more taste
cells (Fig. 1).

The anterior two-thirds of the tongue is supplied
by the lingual nerve, the back of the tongue by the



Fig 1. Highly magnified vertical section of a taste bud from a rabbit. (*D. J. Cunningham, Textbook of Anatomy, 8th ed., Oxford, 1943*)

glossopharyngeal nerve, and the throat and larynx *by branches of the vagus nerve, all utilized for* touch, temperature, and pain sensitivity, as well as for taste The taste fibers from the anterior tongue leave the lingual nerve in a small branch, the chorda tympani, which traverses the eardrum en route to the brain stem. When the chorda tympani is cut or damaged, as in the removal of the tympanum, *taste sensitivity is lost on the anterior two-*thirds of the tongue on the same side. Taste buds degenerate when their nerve supply has been cut, but regenerate if the peripheral nerve fibers regenerate.

The taste fibers from all the sensory nerves from the mouth come together in the solitary tract and its nucleus in the medulla oblongata, in close association with touch and temperature sensory nerve fibers from the tongue The second-order fibers ascend by a pathway, not yet entirely delineated, to a small cluster of cells medially (centrally) placed in the ventral vasal part of the thalamus. Taste fibers, again in association with touch and temperature fibers, project from here to the mouth area of the anterior sensory cerebral cortex. No one part of the cerebrum appears to be exclusively devoted to taste (*see* BRAIN; NERVOUS SYSTEM)

Taste qualities. No simple relation exists between chemical stimuli and taste quality except, perhaps, for the sourness of acids. The taste qualities of inorganic salts are complex, only sodium chloride giving the purely saline taste. Sweet and bitter tastes occur in many different chemical classes

The middorsum (middle top portion) of the tongue surface is insensitive to all tastes. Sensitivity to sweet is greatest at the tip, to sour at the

sides, to bitter at the back; salt sensitivity is relatively more homogeneous around the edges. For many years it was believed that such regional differences were explained by the existence of four different basic types of taste receptors, one for salt, one for sour, one for bitter, and one for sweet, distributed unevenly over the tongue. Electrophysiological records, that is oscillographic tracings,

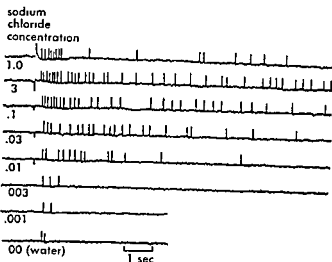


Fig 2. Typical oscillographic record of impluses in a single taste sensory nerve fiber of a rat. (C. *Pfaffmann, Gustatory nerve impulses in rat, cat and rabbit, J. Neurophysiol, 18.432–433, 1955*)

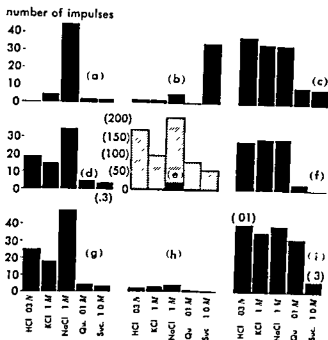


Fig. 3. A bar diagram of the different sensitivity patterns in each of nine single sensory fibers (a–i) from a rat. Solid black indicates number of impulses during first second of discharge to each of five test solutions shown along base: HCl (hydrochloric acid), KCl (potassium chloride), NaCl (sodium chloride), Qu (quinine); Suc. (sucrose). Crosshatching superimposed on fiber e shows relative amount of neural activity in total nerve (sum of all fibers) to same test solutions. (C. *Pfaffmann, Gustatory nerve impulses in rat, cat and rabbit, J. Neurophysiol., 18.432–433, 1955*)

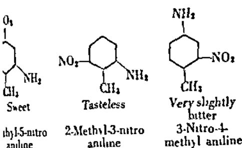
from the individual receptor cells or their asso ated single sensory nerve fibers often show mixe sensitivity; for example, sensitivity to acid an salt, or to acid, salt, and sugar (Fig. 2). There an increase in impulse frequency with increase salt concentration. In some species pure wate strongly stimulates certain taste receptors; that i it causes them to discharge nerve impulses (se BIOPOTENTIALS AND ELECTROPHYSIOLOGY). Tast cells might be said to have a spectrum or wide ban of sensitivities which differ from receptor to recep tor. There is no evidence for the four simple type so that the perception of taste quality depends upo different patterns of sensory nerve discharge en tering the brain (Fig. 3).

Sensitivity. Studies of human sensitivity fo acid by different investigators reveal a wide rang of values, but a median threshold value of 0 000 molar solution (*M*) may be cited as an approxi mate order of magnitude for hydrochloric aci (*see* CONCENTRATION SCALES). Sourness increase with increase in hydrogen-ion concentration, bu weak organic acids are more sour than would b predicted from their hydrogen-ion concentration Increasing carbon chain length in the aliphatic aci series, for example, appears to enhance stimulat ing efficiency

Most salts, except sodium chloride (NaCl) elicit other qualities, like bitter or sour, in addi tion to salty. Low-molecular-weight salts are pre dominantly salty while those of higher molecular weight are bitter, yet the salts of lead and beryl lium are sweet. The molecular weights of lead and beryllium are 207 and 9 respectively. The median human threshold for sodium chloride is approximately 01 *M*, but here again there is a wide range in reported values. Both the anion and cation contribute to saltiness and to stimulating efficiency; for example, in man, the following cation series ranks the contribution to saltiness and efficiency of stimulation in decreasing order: ammonium (NH_4), potassium (K), calcium (Ca), sodium (Na), lithium (Li), and magnesium (Mg). According to one theory, the first step in stimulation occurs when ions bind loosely to the taste receptor surface by a nonenzymatic process, as in the binding of ions by proteins Species differences found in the cation taste series presumably reflect differences in the configuration of the active sites on the receptor surface.

The sweet taste is associated largely with organic compounds such as alcohol, glycols, sugars, and sugar derivatives, with the exception of certain inorganic salts of lead or beryllium. The complex relations between chemical structure and sweetness are not readily explained by present-day systemizations. Sucrose thresholds for man have a median value of .1 *M*, and the synthetic sweetener saccharin is 700 times more dilute at threshold.

Slight changes in spatial arrangement render a molecule tasteless (Fig. 1), but specificity such as this does not appear to involve an enzymatic stimulation process, even though this suggestion has



| 4 Change in taste quality resulting from differ- t spatial arrangements within the molecule. (S. S. vens (ed), *Handbook of Experimental Psychology*, ley, 1951)

en made by some workers. Studies of the differ- tial inhibition of sweet sensitivity by gymnemic id, which does not affect salt or sour, point to a nenzymatic competitive inhibition of specific tes on the receptor surface.

Bitter is elicited by many chemical classes and is ften found in association with sweet and other aste qualities. An increase in molecular weight of organic salts or in the length of a carbon chain f organic molecules may be associated with in- creased bitterness. Typical of substances with the itter taste are the alkaloids such as quinine, caf- eine, and strychnine, which are often toxic. A median threshold value for quinine has been cited at .000008 *M*. Taste blindness, an inherited inabil- ity to taste the bitterness of PTC (phenyl thiocar- bamide) and other substances with the thiocar- bamide >N(C=S)— group does not affect other bitter stimuli without this grouping. About one-third of Caucasians are nontasters of PTC.

The stimulus increment for a noticeable differ-

sitivity increases with temperature rise, salt and quinine sensitivity decrease, but acid sensitivity is unaffected. Taste has no simple temperature coef- ficient as does a simple chemical system.

Adaptation after a solution is flowed continuously over the tongue may lead to a rise in threshold and possibly to complete disappearance of taste sensation. Certain contrast effects may also be noted, so that after exposure to weak acid, distilled water tastes sweet. The bitter or sour taste can be masked by sweetening agents, but there have been few systematic studies of these interactions.

Behavioral effects. The ability of an organism to select nutritious or necessary ingredients of the diet by taste can be demonstrated with the self- selection technique. An animal is given free choice of individual containers with necessary nutrients in pure form. After certain physiological stresses, like glandular imbalances, the selection of the vari- ous nutritive agents often shows compensatory changes. A severely salt-deficient rat (adrenalecto- mized) may show a significant increase in the in- take of sodium chloride sufficient to counteract the usually fatal outcome in the absence of salt-re- placement therapy. Similar effects have been noted in children, but in adults, food habits, cultural con- ditioning, learning, and other complex psychologi- cal factors play a significant role in food accept- ance and may override the physiological factors controlling behavior.

None the less, food palatability, as determined by taste and other sensory effects, has such a pro- found effect on food acceptance that a substantial applied science of flavor technology has developed, particularly in the service of the food industry. *See* SENSE, CHEMICAL; SMELL. [C.P.]

Taurus

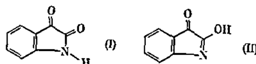
The Bull, in astronomy, is a winter constellation. Taurus is the second sign of the Zodiac. The group contains two notable star clusters, the Hyades and the Pleiades. The Hyades is a V-shaped cluster, the V forming the head of the charging bull, with the fiery bright star Aldebaran in the right eye. This star has long been used in navigation. The long horns of the bull extend northeast to the con- stellation Auriga. Farther west lies the compact, beautiful cluster of six stars, the famous Pleiades, sometimes called the seven sisters, suggesting thereby that one of the stars has faded from naked- eye view. This group in the Bull's shoulder has the shape of a tiny dipper. *See* CONSTELLATION. [C.S.Y.]

Tautomerism

The reversible interconversion of structural iso- mers of organic chemical compounds. Such inter- conversions usually involve transfer of a proton (prototropy), but anionotropic (allylic, Wagner- Meerwein) rearrangements may be reversible and so be classed as tautomeric interconversions.

Lactam-lactim tautomerism. A cyclic system containing the grouping —CONH— is called a lactam, and the isomeric form, —COH=N—, a lactim. These terms have been extended to include the same structures in open-chain compounds when considering the shift of the hydrogen from nitro- gen to oxygen.

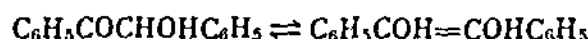
A. von Baeyer first recognized that isatin (I) ap- pears to react in either the lactam (I) or the lac- tim (II) structure. Thus, a precedent for proving



the existence of lactim-lactam tautomerism was established; chemical behavior as inferred from the structure of a reaction product. More recently spectroscopic techniques have been judged more reliable, and it is often possible to determine from the absorption spectrum whether a given substance has either one structure or both.

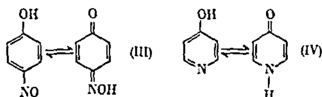
Keto-enol tautomerism. The molecular grouping, $—COCH<$, may in certain substances exist partly or wholly as $—COH{=}C<$. The former constitutes the keto form and the latter the enol form. Kurt Meyer first studied the keto, $CH_3COCH_2CO_2C_2H_5$, and enol, $CH_3COH{=}CHCO_2C_2H_5$, forms of ethyl acetoacetate, and which he recognized respectively by reactions specific for the carbonyl group and the carbon-carbon double bond. Both forms may be obtained in relatively pure condition: the former by freezing it out of the mixture and the latter by slowly distilling the mixture in quartz apparatus. However, each is slowly converted into the equilibrium mixture of the two. Extensive chemical and spectroscopic studies have shown that the enol content of such an equilibrium mixture is a function of the physical state of any given substance. The gas phase or solution in a nonpolar solvent (hexane) favors the enol form, whereas more polar solvents (chloroform, alcohols) repress its formation.

The existence of an enol in an acyclic system requires that a second carbonyl group (or its equivalent, for example, $>C{=}N—$) be attached to the same $—CH<$ as an aldehyde or ketone carbonyl. Thus, ethyl acetoacetate tautomerizes demonstrably, but ethyl malonate, $C_2H_5O_2CCH_2CO_2C_2H_5$, does not. Occasionally an enol form exists, these requirements notwithstanding: for example, ethyl pyruvate is partially enolized ($CH_3COCO_2C_2H_5 \rightleftharpoons CH_2{=}COHCO_2C_2H_5$); and α-hydroxy ketones (or aldehydes) exhibit the characteristics of the tautomeric enediols, for example, benzoin

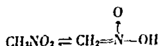


$$C_6H_5COCHOHC_6H_5 \rightleftharpoons C_6H_5COH{=}COHC_6H_5$$

Where the enol form includes an aromatic ring such as phenol, the existence of the keto form is often not demonstrable, although in some substances such as 4-nitrosophenol (III) and 4-hydroxypyridine (IV), there may be either chemical or spectroscopic evidence for both forms.



Closely related to keto-enol tautomerism is the prototropic interconversion of nitro and aci forms of aliphatic nitro compounds such as nitromethane



$$CH_3NO_2 \rightleftharpoons CH_2{=}\overset{O\uparrow}{N}—OH$$

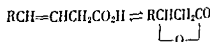
Ring-chain tautomerism. The possibility that an acyclic hydroxyaldehyde may exist in equilibrium with its cyclic hemiacetal was first recognized by Emil Fischer. The failure of glucose to form a normal acetal with an alcohol and the surprising production of two isomeric glucosides instead led to the postulate that carbohydrates exist principally as inner or cyclic hemiacetals in equilibrium with only enough free aldehyde to permit typical aldehyde reactions with reagents which either oxidize the carbonyl group or form derivatives that effectively remove it from the equilibrium:



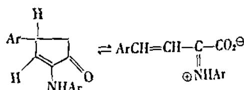
The glycosides are formed by the elimination of water between the hydroxyl derived by hemiacetal formation and an alcohol, with two structures being possible, and the hydroxyl lying above or below the hetero ring. *See* GLUCOSE.

In general, tautomeric forms will exist in substances possessing functional groups which can interact additively and which are so placed that intramolecular reaction will lead to a stable cyclic system. The cyclic form will usually predominate (especially if it contains five or six members).

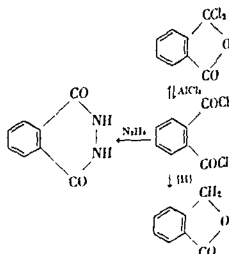
R. P. Linstead has shown that certain alkenic acids are tautomeric with their lactones



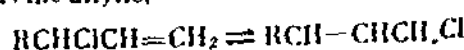
and has called this lacto-enoic tautomerism. More recently lacto-enoic tautomerism not involving a prototropic shift has been observed:



And still another type of ring-chain tautomerism not involving a prototropic shift is demonstrable by the reactions of phthaloyl chloride:

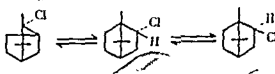


The latter type of ring-chain tautomerism is closely related to anionotropic rearrangements such as the allylic,



$$RCHClCH{=}CH_2 \rightleftharpoons RCH{-}CHCH_2Cl$$

and Wagner-Meerwein.



which may thus be considered as examples of anionotropic tautomerism. *See* ISOMERISM, MOLECULAR. [W R V.]

Taxis

A mechanism of orientation by means of which an animal moves in a direction related to a source of stimulation. There exists a widely accepted terminology in which the nature of the stimulus is indicated by a prefix such as phototaxis, chemotaxis, geotaxis (gravity), thigmotaxis (contact), rheotaxis (water-current), and anemotaxis (air-current). The directions toward or away from the stimulus are expressed as positive or negative, respectively. Finally, the sensory and locomotory mechanisms by means of which the orientation is achieved are denoted by a second type of prefix forming a compound noun with taxis. Positive phototropotaxis thus describes a mechanism by means of which an animal carries out a directed movement toward a source of light along a path which permits the animal's paired eyes to receive equal intensities of light throughout the movement. The following are examples of various types of taxes.

Klinotaxis. A well analyzed case of this type of taxis is the way in which a fly maggot moves away from the light immediately before pupating in a dark and sheltered place. When such a maggot is exposed to a horizontal beam of light above the substrate, it moves along a fairly straight path away from the light in the direction of the beam. In doing so it waves its front end from side to side exposing alternately the left and right lateral aspects of its front end to the light shining from behind. As long as the light intensity falling on the light-sensitive surfaces on either side remains equal in subsequent exposures, the animal will follow a straight path away from the light source. This may be explained by the hypothesis that the extent of the swing toward one side is a function of the light intensity falling on the other. If the animal starts its course from a position at an angle with the light beam, the differential light intensities falling on its anterior flanks automatically steer it into a path curving into line with the beam. The prefix "klino" in this case denotes that "exploratory" side-

has now become restricted to the description of orientation phenomena in sessile organisms. The essential point is the unwavering turn of the organism into the stimulus direction by means of innate reflex mechanisms linking bilaterally symmetrical receptors with the organs of locomotion. Paths toward or away from a single source are straight. In the case of two or more sources, they run along the resultant of incident intensities. After unilateral receptor loss, tropotactic steering leads, in a uniform field of illumination, to continued "circus movement" away from or toward the injured side in the case of positive or negative taxis respectively.

Telotaxis. Whereas bilateral intensity balance on receptors is essential in klinotaxis and tropotaxis, orientation in telotaxis occurs, as it were, by aiming one or the other of two bilaterally symmetrical receptors toward the stimulus. Of a number of simultaneously offered stimuli, all but one may be "ignored" at any given instant by means of built in switch mechanisms based on inhibition or block. Unilateral sense organ loss does not lead to circus movement.

A special case of telotaxis is the light compass reaction. In this the animal moves at a temporarily fixed angle with respect to the stimulus direction. The angle can be changed at will. The term pharotaxis has been suggested to describe an orientation toward a progressively changing direction of stimulus, such as the change of distribution of polarization of the light of the sky during the course of the day (navigation by honeybees and other arthropods).

The description of taxes given so far is based on concepts formulated by Alfred Kuhn and after him by G. Fraenkel and D. Gunn. They apply largely to mechanisms of orientation found in small-brain animals which rely in their basic orientation on a relatively small number of innate responses. These are characterized by their stereotyped and rather inflexible nature. In recent years O. Koehler and with him a number of students of animal behavior have attempted to make the taxis concept part of the more complex and plastic behavior of vertebrates, including man. Thus Koehler's definition of a taxis is as follows: every purely reflexive and every voluntary act begins with certain postures and movements which orient the animal's head, limbs, and body with reference to the direction of the eliciting stimulus or of the final goal of the movement. These intentional postures or movements of postural adjustment are to be considered the taxis components of the animal's act. Even in man, the turning of eyes or head into the direction of an object of interest or desire is held to fall within this definition. [O.L.]

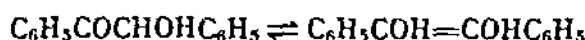
Bibliography: G. S. Fraenkel and D. L. Gunn, *The Orientation of Animals*, 1940; O. Koehler, Die analyse der taxisanteile instinktartigen verhaltens, *Physiological Mechanisms in Animal Behaviour*, Soc. Exptl. Biol., Symposium 4:269–304, 1950.

Taxodonta

A subclass of pelecypod mollusks in which the hinge is of the taxodont type; that is, the dentition is a series of similar alternating teeth and sockets along the hinge margin (see illustration). Geologically, these marine mollusks range from the Ordovician to the Recent. The adductor muscles are approximately of equal size, and gills are present for both respiratory and feeding purposes. *Nucula* is one of the best known modern representatives with 7 species. It occurs at depths of 10–1000 fathoms where it moves on the surface of sandy or

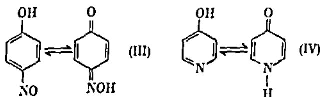
Keto-enol tautomerism. The molecular grouping, —COCH<, may in certain substances exist partly or wholly as —COH=C<. The former constitutes the keto form and the latter the enol form. Kurt Meyer first studied the keto, $CH_3COCH_2CO_2C_2H_5$, and enol, $CH_3COH{=}CHCO_2C_2H_5$, forms of ethyl acetoacetate, and which he recognized respectively by reactions specific for the carbonyl group and the carbon-carbon double bond. Both forms may be obtained in relatively pure condition: the former by freezing it out of the mixture and the latter by slowly distilling the mixture in quartz apparatus. However, each is slowly converted into the equilibrium mixture of the two. Extensive chemical and spectroscopic studies have shown that the enol content of such an equilibrium mixture is a function of the physical state of any given substance. The gas phase or solution in a nonpolar solvent (hexane) favors the enol form, whereas more polar solvents (chloroform, alcohols) repress its formation.

The existence of an enol in an acyclic system requires that a second carbonyl group (or its equivalent, for example, >C=N—) be attached to the same —CH< as an aldehyde or ketone carbonyl. Thus, ethyl acetoacetate tautomerizes demonstrably, but ethyl malonate, $C_2H_5O_2CCH_2CO_2C_2H_5$, does not. Occasionally an enol form exists, these requirements notwithstanding: for example, ethyl pyruvate is partially enolized ($CH_3COCO_2C_2H_5 \rightleftharpoons CH_2{=}COHCO_2C_2H_5$); and α-hydroxy ketones (or aldehydes) exhibit the characteristics of the tautomeric enediols, for example, benzoin

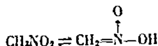


$$C_6H_5COCHOHC_6H_5 \rightleftharpoons C_6H_5COH{=}COHC_6H_5$$

Where the enol form includes an aromatic ring such as phenol, the existence of the keto form is often not demonstrable, although in some substances such as 4-nitrosophenol (III) and 4-hydroxypyridine (IV), there may be either chemical or spectroscopic evidence for both forms

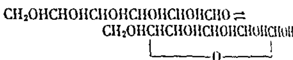


Closely related to keto-enol tautomerism is the prototropic interconversion of nitro and aci forms of aliphatic nitro compounds such as nitromethane



$$CH_3NO_2 \rightleftharpoons CH_2{=}\overset{\overset{O}{\uparrow}}{N}{-}OH$$

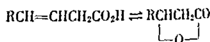
Ring-chain tautomerism. The possibility that an acyclic hydroxyaldehyde may exist in equilibrium with its cyclic hemiacetal was first recognized by Emil Fischer. The failure of glucose to form a normal acetal with an alcohol and the surprising production of two isomeric glucosides instead led to the postulate that carbohydrates exist principally as inner or cyclic hemiacetals in equilibrium with only enough free aldehyde to permit typical aldehyde reactions with reagents which either oxidize the carbonyl group or form derivatives that effectively remove it from the equilibrium:



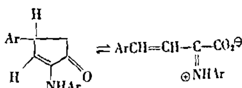
The glycosides are formed by the elimination of water between the hydroxyl derived by hemiacetal formation and an alcohol, with two structures being possible, and the hydroxyl lying above or below the hetero ring. See GLUCOSE.

In general, tautomeric forms will exist in substances possessing functional groups which can interact additively and which are so placed that intramolecular reaction will lead to a stable cyclic system. The cyclic form will usually predominate (especially if it contains five or six members).

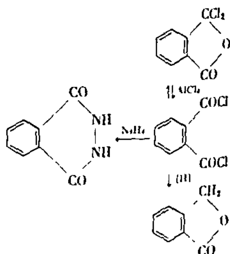
R. P. Linstead has shown that certain alkenic acids are tautomeric with their lactones



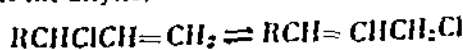
and has called this lacto-enoic tautomerism. More recently lacto-enoic tautomerism not involving a prototropic shift has been observed:



And still another type of ring-chain tautomerism not involving a prototropic shift is demonstrable by the reactions of phthaloyl chloride:

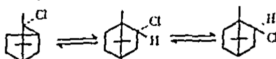


The latter type of ring-chain tautomerism is closely related to anionotropic rearrangements such as the allylic,



$$RCHClCH{=}CH_2 \rightleftharpoons RCH{=}CHCH_2Cl$$

and Wagner-Meerwein,





Tea plant, *Thea sinensis*. (From L. H. Bailey, ed., *The Standard Cyclopedia of Horticulture*, vol 3, Macmillan, 1935)

of new shoots, the source of commercial tea The leaves are picked by hand and dried. Quick drying by artificial heat produces green tea Other processes yield black and oolong teas

Tea is a crop of hot, temperate regions. Some 1000 varieties are listed. India, Ceylon, China, Formosa, Java, and Japan are the leading tea-producing countries, with China contributing about one-half of the world's supply. *See* PARIETALES.

[P.D S.]

TEA PRODUCTION

Tea production technology depends on whether black, green, or semifermented tea is to be manufactured The characteristic appearance and flavor of each arise more from different processing and manufacturing methods than from the type of leaf used

Black tea. This tea requires quick transportation of the green leaf from garden to factory, where it is carefully spread on withering tats The leaf loses moisture and becomes flaccid within 24 hours. Circulation of conditioned air during withering shortens the time substantially. The next process, called rolling, ruptures the leaf cells and releases their juices. This is accomplished in a brass roller box, equipped with adjustable pressure cap and an open base. After filling with withered leaf, it is rotated eccentrically over a brass table fitted with curved battens. During a 30-min period, the leaf cells rupture due to imposed pressure and the rubbing action. A second rolling period causes twisting and coating of the leaf with juices, leading to balling The leaf changes to a bright copper color, and a characteristic tea aroma develops A roll breaker is subsequently employed to disintegrate the balls for fermentation

In fermentation enzyme action and oxidation occur. The rolled leaf is spread to a 2- or 3-in. depth on fermenting beds and held under high humidity at 75–80°F. Since the body, or strength, of the tea depends on this fermentation, careful attention to its progress is required. Approximately 3½ hours are needed for the rolling and fermentation operations.

Firing or drying of fermented tea leaf conventionally is carried out by passing it through heated ovenlike chambers. Temperatures of 170–180°F arrest fermentation and develop the familiar blackish color. Drying is practically completed in one pass, but further heat is applied to produce the "case hardening," that protects quality. The leaf is sieved, graded and packed into aluminum foil-lined chests, containing about 100 lb of finished tea.

Green tea. Green tea manufacture requires that the plucked leaves be steamed as quickly as possible. Such processing at 160°F makes them soft and pliable and, by inactivating the natural enzymes, prevents fermentation. After steaming, the leaf is alternately rolled and dried until it becomes too stiff for further manipulation. For export, green tea is refired in pans with mechanical stirring to produce a luster.

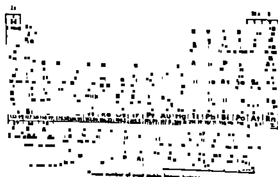
Oolong teas. Oolong teas are midway between black and green teas in that they are semifermented. After a short sun-withering in the garden, the leaf is gently rolled in the plucker's hands, whereupon a slight fermentation is initiated. After a short period this leaf is sent to the factory to be fired and packed for shipment.

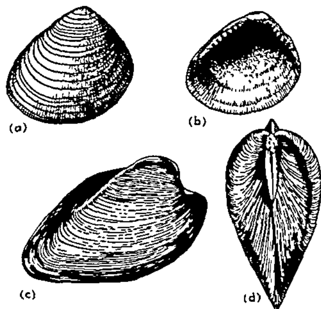
Instant tea. This type of tea is obtained by spray-drying of a black tea extract, with or without the admixture of maltodextrins The technology is patterned closely after that followed in manufacturing of instant coffee. Concentrated tea extract also may be combined with a heavy sugar syrup and marketed as a liquid. Solubility of instant tea powder can be varied widely by manipulations in processing *See* FOOD ENGINEERING. [J H.N.]

Bibliography: C. R. Harler, *The Culture and Marketing of Tea*, 2d ed., 1956

Technetium

Element number 43, technetium, Tc, was discovered by C. Perrier and E. G. Segré in 1937 as a result of cyclotron bombardment of molybdenum by deuterons. It was the first element to be made artificially, hence the name technetium. It also occurs as a major constituent in the fission products resulting from the operation of a nuclear reactor





Taxodonta (a) Exterior of right valve of *Nucula defuniak*, a Miocene species. (b) Interior of left valve of *N. defuniak* (c) Exterior of right valve and (d) dorsal view of complete shell of *Cyrtodonta hindi* from the Ordovician. (*From R. R. Shrock and W H Twenhofel, Principles of Invertebrate Paleontology, 2d ed., McGraw-Hill, 1953*)

silty bottoms. The shell is small and ranges from 6 to 15 mm. See PELECYPODA. [C.B C.]

Taxonomic categories

One of a hierarchy of levels, or taxa (singular, taxon), in which organisms are classified to indicate various degrees of relationship. Two major kinds of categories are utilized, those which pertain to groupings of individuals and populations of individuals (specific and infraspecific categories) and those which involve groupings of species (higher categories) In this taxonomic hierarchy, the category species is the most important, and is generally thought of as a distinctive interbreeding or potentially interbreeding population of relatively similar individuals which differs from other such populations in one or more recognizable structural or physiological characteristics (*see* SPECIES CONCEPT).

When individuals of different species are compared with one another they are found to have varying degrees of similarity, and since, generally speaking, degree of similarity is indicative of degree of relationship, they can be arranged in an ascending series of categories which express these degrees of similarity. Groups of closely related species are placed together in a genus, related genera in a family, related families in an order, related orders in a class, related classes in a phylum, and related phyla in a kingdom. Each of these categories is more comprehensive than the preceding. At a given level, each taxon is separated from the most nearly related taxa by a gap in similarity of characteristics indicating a gap in degree of relationship. The components of each taxon are presumed to have a common phylogenetic origin and most, but not all, taxa contain more than one unit in the next lower category.

The respective position of two different animals in this system may be expressed as follows:

	Lion	*Housefly*
Kingdom	Animalia	Animalia
Phylum	Chordata	Arthropoda
Class	Mammalia	Insecta
Order	Carnivora	Diptera
Family	Felidae	Muscidae
Genus	*Felis*	*Musca*
Species	*leo*	*domestica*

These seven taxonomic categories are the minimum used to classify most animals. In large groups with a long or complex evolutionary history, more taxa are evident and more categories are utilized. Those which are generally accepted are as follows:

Kingdom
  Phylum
    Subphylum
      Superclass
        Class
          Subclass
            Cohort
              Superorder
                Order
                  Suborder
                    Superfamily
                      Family
                        Subfamily
                          Tribe
                            Genus
                              Subgenus
                                Species
                                  Subspecies

Standardized endings are utilized for superfamilies, families, subfamilies, and tribes (*see* ZOOLOGICAL NOMENCLATURE). Names for categories above the superfamily are not standardized and vary from one group of animals to another. *See* PLANT CLASSIFICATION. [E G LI.]

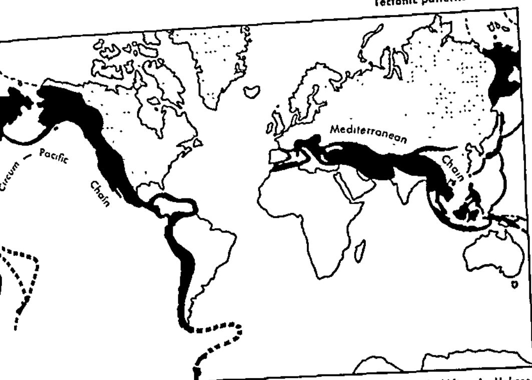
Bibliography: E Mayr, E. G. Linsley, and R. L. Usinger, *Methods and Principles of Systematic Zoology*, 1953.

Taxonomy

Taxonomy is concerned with the identification, naming, and classification of living organisms. It involves the collection, recording, and preservation of specimens and the making of identification keys and manuals, so that each plant or animal will be accurately described, correctly placed in the taxonomic scheme, and will bear a single and universally recognized scientific name. Taxonomy also includes the rules of classification. *See* ANIMAL SYSTEMATICS; PLANT TAXONOMY; *see also* BACTERIA, TAXONOMY OF; FUNGI; PALEOBOTANY, PALEONTOLOGY; PROTOPHYTA, SCHIZOMYCETES [P D.S.]

Tea

The popular caffeine beverage made from the leaves of the tea plant, *Thea sinensis*, a member of the tea family, Theaceae. The plant is a small tree, but in cultivation constant pruning makes it a shrub 3-4 ft tall. Pruning promotes development



1 Late Mesozoic and Cenozoic mountain chains the world (black areas) Gondwanaland (shaded areas) and Laurasia (dot pattern). (*After A. Holmes, Principles of Physical Geography, Ronald, 1945*)

rth's crust, and considers the cause or processes deformation to the bottom of the crust, 35–70 lometers (km) deep, and into the mantle, some 0 km deep. For a discussion of crustal move-ents and the origin of tectonic forces *see* DIAS-ROPHISM, OROGENY, TECTONOPHYSICS

The master divisions of the earth's crust are the ocean basins and the continents. The continents may be divided into stable regions (generally of low relief) and unstable or mobile regions (the mountain systems). Each of these continental divi-sions, of course, has a number of subdivisions For instance, stable regions usually have a portion which has had a history dominantly of uplift, and consequently has been stripped of its younger strata, exposing the older (Precambrian) rocks over wide areas. The Canadian shield of North America is an example. Other parts of stable re-gions have subsided dominantly over long periods of time and were, consequently, sites of inland seas. Such areas, like the Michigan basin, are cov-

... arches or domes.

The continents of South America, Africa, Aus-tralia and Antarctica have been referred to col-lectively as Gondwanaland because they have cer-tain unifying geologic characteristics, and those north of the Mediterranean belt and east of the North American cordillera as Laurasia (Fig 1) Gondwanaland and Laurasia, aside from the mod-ern cordilleras, are composed of stable shields and ancient mountain belts, now reduced to hills or modest mountains by long erosion. As shown in Fig. 1, the youthful and high mountains of Gond-wanaland and Laurasia are cast in two majestic chains. One is the east-west Mediterranean belt that stretches from Spain and Morocco to Malaya and the Dutch East Indies The other is the Circum-Pacific belt that embraces the cordilleras of South and North America, the eastern Asiatic island archipelagos, and the New Zealand–New Guinea systems. This latter highly complex region, the Dutch East Indies, is the junction of the great Mediterranean and Circum-Pacific belts.

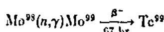
The major tectonic features of the world are dis-... literranean ... dillera, is-

MEDITERRANEAN CHAIN

... Indies arcuate complex. The Alpine Cordillera is composed of the Atlas, Pyrenees, Alps, Apennines, and Carpathians. Of these, the Alps are undoubt-edly the most studied mountains in the world. The following discussion is based largely on the classic description by A Holmes in 1945.

The Alps. Geologically the Alps are divided into the Western Alps and Eastern Alps The Western Alps extend northward in a broad arc from the Gulf of Genoa to Lakes Geneva and Constance The Eastern Alps continue in a gentler curve toward Vienna Beyond Vienna, the Carpathians represent an eastward continuation of the Alpine structure South of the high Alps the ranges of northern Italy extend eastward and merge into the Dinaric Alps It is in the Western Alps, and particularly in Swit-

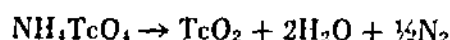
or, alternatively, by action of neutrons on Mo^{98} where the reaction is



$$Mo^{98}(n,\gamma)Mo^{99} \xrightarrow[67\ hr]{\beta^-} Tc^{99}$$

The isotope Tc^{99} is most suitable for chemical investigation because of its long half-life, 2×10^5 years. The chemistry of technetium is very similar to that of rhenium and corresponding compounds have been prepared in many cases. The metal can be prepared by reduction of the sulfide with hydrogen at temperatures of 1000–1100°C, and its crystal structure has been found to be isomorphous with that of rhenium, osmium, and ruthenium.

Technetium metal reacts with oxygen at elevated temperatures to form the volatile oxide Tc_2O_7 which is analogous to Re_2O_7. Another oxide, TcO_2, is formed by the decomposition of NH_4TcO_4 at elevated temperatures in vacuum according to the equation



$$NH_4TcO_4 \rightarrow TcO_2 + 2H_2O + \tfrac{1}{2}N_2$$

Reactions which produce the compounds $AgTcO_4$, $KTcO_4$, NH_4TcO_4, K_2TcCl_6, and TcS_2 are analogous to those used to form the corresponding rhenium compounds *See* NUCLEAR REACTION; RHENIUM, TRANSITION ELEMENTS. [S.F.]

Bibliography · S Tribalat, *Rhénium et Technétium*, 1957.

Technology

Systematic knowledge and action, usually of industrial processes but applicable to any recurrent activity. Technology is closely related to science and to engineering Science deals with man's understanding of the real world about him—the inherent properties of space, matter, energy, and their interactions (*see* SCIENCE) Engineering is the application of objective knowledge to the creation of plans, designs, and means for achieving desired objectives (*see* ENGINEERING) Technology deals with the tools and techniques for carrying out the plans

For example, certain manufactured parts may need to be thoroughly clean The technological approach is to use more detergent and softener in the wash water, to use more wash cycles, to rinse and rerinse, and to blow the parts dry with a stronger, warmer, air blast. Often such refinements provide an adequate action However, if they do not suffice the basic technique may need to be changed Thus, in this example, science might contribute the knowledge that ultrasonically produced cavitation counteracts surface tension between immiscible liquids and adhesion between clinging dirt and the surface to be cleaned, and thereby produces emulsions. Engineering could then plan an ultrasonic generator and a conveyor to carry the parts through a bath tank where the ultrasonic energy could clean them. The scientist may use ultrasonic techniques to determine properties of materials. The engineer may design other types of devices that employ ultrasonics to perform other functions. These specialists enlarge their knowledge of ultrasonics and their skill in using this technique not for its own sake but rather for its value in their work. The technologist is the specialist who carries out the technique for the purpose of accomplishing a specified function. He extends his knowledge and skill of ultrasonic cleaning by refinement and perfection of the technique for use on various materials soiled in different ways. Technological advances improve and extend the application to cleaning other parts under other conditions. *See* MECHANICAL ENGINEERING. [R.N.SH.]

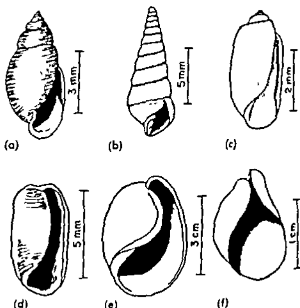
Tectibranchia

An order in the subclass Opisthobranchia containing the sea hares and the bubble shells. The shell may be present, rudimentary, or absent. When the shell is

the mant

families

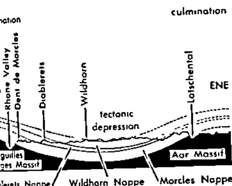
ble shells have a rather thin shell with a wide, capacious aperture, the shell being enveloped by the mantle when the animal is actively crawling (see illustration). The sea hares are usually without a shell and can crawl over the sea floor, and many can swim, using a remarkable undulatory motion Many sea hares emit a purple dye when disturbed This dye is soluble in sea water and should not be confused with the purple dye emitted by certain other species of sea snails, the source of Tyrian purple of ancient Syria They are predatory and feed mainly on small crustaceans [W.J.C.]



Tectibranchia (a) *Acteon* (b) *Pyramidella* (c) *Retusa* (d) *Cylichna* (e) *Bulla* (f) *Haminoea* (From A. M. Keen and J C Pearson, *Illustrated Key to West North American Gastropod Genera*, Stanford Univ. Press, 1958)

Tectonic patterns

The arrangement of the large structural units of the earth's crust, such as mountain systems, shields or stable areas, basins, arches, and volcanic archipelagos. Tectonic geology pertains to the constitution and deformation of these large parts of the



3 Section showing nappes of High Calcareous s exposed in tectonic depression between culminas. Nappes advanced at right angles to the section direction away from the observer. (*Adapted from Holmes, Principles of Physical Geology, Ronald,* 45)

nes). These rocks represent the floor and later posits of the geosyncline (Fig. 5).

Within the Pennides, the Simplon Nappes Monte Leone, Lebendum, and Antigorio) are the wer members of the series. Of the higher nappes, e Great St. Bernard Nappe overrides the Simplon appes; the Monte Rosa Nappe plunges in back f the St. Bernard Nappe; and the Dent Blanche appe was thrust forward over all the others but as since been removed by erosion (Fig 2).

To the east the Pennine Nappes continue at a ower level as the Lepontine Alps (Austrides of

High Tauern, where the two upper members of the series have been recognized (Fig. 4)

Zone of roots The "roots" of the Pennine Nappes lie in a long narrow zone near the Italian frontier. Here the nappes turn vertically down and appear to be rooted in the ground On the south side and in contact with the Pennine root zone lie the roots of the Austrides, or nappes of the Eastern Alps. Still farther south are the roots of the Dinarides, or nappes of the Dinaric Alps To the east the root zone gradually widens with the introduction of the Austride and Dinaride structural elements

The originally great width of the geosyncline is implied in the sedimentary strata of the nappes Horizontal compression culminated in Miocene times and during its progress the rocks must have been unusually plastic. Evidence of lubrication by hot migrating fluids is furnished by the prevalence of migmatites, with swirling structures produced by flowage, in the deepest parts of the roots

Apennines. The Apennines traverse peninsular Italy and according to one theory form the outer zone of a mountain system, the core of which is now submerged beneath the Tyrrhenian Sea From the core area, which was a linear element of uplift, great landslides, mostly submarine, moved huge masses of

Apennines.

Carpathians. The Carpathians represent a continuation of Alpine structure eastward, but with variations An internal zone is composed of pre-Carboniferous crystalline rocks and a nearly complete Mesozoic sequence. This was deformed in Late Cretaceous and Early Tertiary time. The outer zone consists of Cretaceous and Early Tertiary sediments which were cast into overturned folds and thrust sheets during Miocene time. The movement of thrust sheets was toward the north

Pyrenees. The Pyrenees are a belt of sharp folds and overthrusts involving principally massive Mesozoic limestones The Aquataine Basin to the north is the site of thick Cretaceous and Tertiary sediments, which have been derived from the rising Pyrenees Many Alpine characteristics have been noted.

Middle East Cordillera. Under this term the mountain complex of Turkey, the Caucasus of Georgia, the mountain complex of Iran (also called

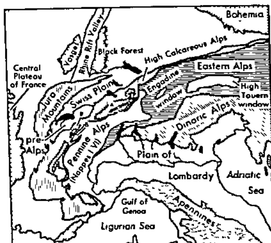


Fig 4 Tectonic map of the Alps A, Aar Massif, B, Mt Blanc, G, St Gotthard Massif, R, Aiguilles Rouges. (*Adapted from A. Holmes, Principles of Physical Geology, Ronald, 1945*)

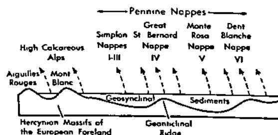


Fig 5 Diagram showing stage in the development of the Tethys geosyncline and its northern shores, to illustrate the environments from which the nappes of the Western Alps were driven. (*Adapted from A Holmes, Principles of Physical Geology, Ronald, 1945*)

zerland, that intensive studies have revealed the key to the general structure of the Alps.

Nappes are the essential feature in the Western and Central Alps. These huge overthrust masses were pushed northward many miles, and some southward. They consist of strata which were highly folded during the movement, much like the contortions of a viscous liquid after flowage (Fig. 2).

Alpine rivers have deeply dissected the nappes and thus have exposed their internal structure along many steep-walled valleys. As the nappes are traced along the trend of ranges the strata undulate in a succession of broad culminations and depressions (Fig. 3). In the depressions the uppermost strata are preserved and the structure is seen in the present mountain peaks. In the culminations, where the higher strata have been removed by erosion, the structure of the lower strata may be seen in the deep valleys. The whole complicated structure can be visualized by taking a series of sections across the culminations and depressions.

The structures seen in the Western Alps are covered by higher nappes in the Eastern Alps where they are seen only in local areas in which erosion has removed the overlying nappes and opened windows, such as the Engadine and High Tauern. The chief subdivisions in the Western Alps are shown in Fig. 4.

Jura Mountains The folds of the Juras form an arcuate bundle of hills between the central plateau of France on the one side and the Vosges and Black Forest on the other The disrupted strata represent the foreland of the Alpine movements. The outermost zone is a tableland broken by faults into irregular strips and blocks On the inner side the strata are thrown into a series of anticlines and synclines; some of the anticlines form the actual hills.

Swiss Plain. This broad lowland lies between the Juras and the High Calcareous Alps It is filled with soft Tertiary sediments, called molasse, derived from denudation of the rising Alps. The plain is interrupted by foothills protruding from the higher Alps, known as the Pre-Alps

Pre-Alps. The Pre-Alps extend between Lake Thun and the River Arve. The strata of these isolated nappes are different from those found in the Juras and High Calcareous Alps and are completely foreign to the district in which they came to rest. The rock sheets are much folded and sliced by minor thrusts and have been driven far from their source. Exactly where they came from remains an unsolved problem of Alpine tectonics. They may represent remanents of nappes that formerly covered the Western Alps as a continuation of the upper nappes of the Eastern Alps. It is possible that these isolated nappes reached their present position by down-sliding similar to a gigantic landslip.

High Calcareous Alps. This high range of rugged mountains (including such peaks as the Jungfrau, 13,669 ft) is made up of a series of clean-cut overthrusts. Locally, many of the overthrust nappes are intensely folded. The strata of the nappes are composed of sediments deposited along the northern margin of the Alpine geosyncline. The High Alps zone includes the Bernese Oberland with its snowfields and glaciers. See GEOSYNCLINE.

Hercynian Massifs. The Hercynian Massifs include two arcuate chains or zones of isolated blocks that resisted the northward advance of the Alpine folds Included in the outer group are the central plateau of France, and the Vosges-Black Forest and Bohemian massifs (Fig 4). The massifs of the inner group occupy positions within the High Alps, or "zone of Mont Blanc." These deeply eroded massifs form the jagged skyline of the Aiguilles Rouges and the adjoining Mont Blanc (15,732 ft). Both the Aiguilles Rouges and Mont Blanc massifs emerge along the crests of a great culmination. To the northeast the Hercynian foundation is covered by the nappes of the High Calcareous Alps (Fig. 3) The nappes at this point lie in a tectonic depression beyond which Hercynian elements emerge in another culmination as the Aar and St Gotthard massifs

Pennine Nappes The Pennine Alps comprise a lofty region of pyramidal peaks, including the Matterhorn (14,705 ft) and Monte Rosa (15,215 ft) Structurally, the Pennine Nappes, or Pennides, are composed of a series of six great recumbent folds (nappes) which were squeezed out of the Alpine (Tethys) geosyncline. Each nappe has a core of older rocks, mainly gneisses and an envelope of metamorphosed rocks (schistes lustrés) and newer sedimentary rocks (crystalline lime-

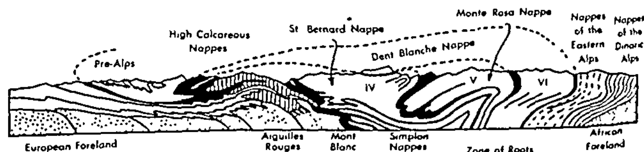


Fig. 2. Cross section of the Western Alps. Dashed lines show probable past position of eroded strata. (Adapted from A. Holmes, Principles of Physical Geology, Ronald, 1945)

formed Most of the major valleys are of structural origin rather than erosional.

Coast Ranges. The Coast Ranges of California, Oregon, and Washington are parts of a young mountain belt welded seemingly by compressional forces on the edge of the continent. They involve the western margin of the older Nevadan belt and contain in places great thicknesses of Cretaceous and Cenozoic strata which accumulated after the batholithic intrusions. The San Andreas fault is one of a related system along which the oceanward block of crust has moved horizontally northwestward for many miles It is still very active.

Mexico. The Nevadan belt extends southward the length of Baja California and probably through the Sierra Madre del Sur The Gulf of California appears to be a downfaulted zone with the San Andreas fault system extending into it from the north The Sierra Madre Occidental is a vast volcanic field over an underlying structural complex similar to the Rocky Mountains and Great Basin system of the United States The Sierra Madre Oriental is a continuation of Rocky Mountain structure into east-central Mexico from west Texas and New Mexico.

South American Cordillera. This cordillera, although supporting much higher peaks, is considerably narrower than the North American Cordil-

tance from Columbia to Tierra del Fuego Inland and flanking the Nevadan belt is a fold and thrust belt of younger age, somewhat like the Rocky Mountain system but generally narrower. In places the two belts are separated by long narrow downdropped fault blocks, but in other places the faulting transgresses both belts Associated with the faulting, if not everywhere geographically then at the same time (Tertiary), are large outpourings of volcanic rock which extensively cover the batholithic and fold belts The highest peaks of Peru, Bolivia, and Chile are volcanic cones built above the general terrane Along the coast of Peru and Ecuador coast ranges have been added to the Nevadan belt in more recent geologic time

Asiatic island arc complex. The Circum-Pacific mountain chain of late Mesozoic and Cenozoic age is represented by an imposing array of island arcs in the western Pacific very little of the eastern margin of the mainland of Asia is involved in modern mountain building. Starting with the island arc of the Aleutians the belt is traced through Kamchatka and the Kuriles to the Japanese archipelago, and then southward in two major spurs one along the Ryukyus (Nansei Shoto) to Formosa and the Philippines, and thence to the Dutch East Indies complex; the other through the Bonine and Mariana arcs and Yap and Palau to Halmahera of the East Indies. This is a region of vigorous mountain growth at the present. Its characteristics are discussed under island arcs.

The wide belt of mountains occupying nearly all of Alaska, except the Arctic Coastal Plain, extends into Siberia. Recent U.S.S.R. maps show that the belt curves southward around a small central shield and projects under the Okhotsk Sea. Evidently, this main and somewhat older cordillera lies submerged under the Okhotsk and Japanese Seas between the mainland and the island archipelagos.

OLDER CORDILLERAS

Major mountain systems older than those heretofore described occur extensively in North America, Eurasia, Australia, and Africa. The Appalachian Highlands include rather ancient systems in Newfoundland, the Maritime Provinces, and New England where the times of mountain building were Late Ordovician (the Taconic orogeny) and Late Devonian (the Acadian orogeny). The classical Appalachians of the eastern United States, whose parallel flat-topped ridges are erosional remnants of folds and thrust sheets, were formed in Pennsylvanian and Permian time (the Appalachian orogeny). The crystalline piedmont of Virginia, the Carolinas, and Georgia seems to be a continuation of the Taconic and Acadian belts of New England.

In Europe, most of Norway is a mountain belt of Late Silurian age (the Caledonian orogeny). It projects southwestward across the central and northern parts of the British Isles Discordantly superimposed on the Caledonian belt and lying across southern England is a later belt of Pennsylvanian and Permian age, the Hercynian. This spreads through most of France and parts of Germany north of the Alps and Juras

The foundations of Brittany, the Vosges, the Ardennes, and the central plateau in France are Hercynian The same is true of the Black Forest, the Harz Mountains, and the Bohemian massif of Germany Most of Spain is Hercynian mountain structure and basement elements of the younger Pyrenees and Alps, such as the Mont Blanc and Aiguilles Rouges massifs, reveal Hercynian structures

The depressed regions between these various remnants of late Paleozoic mountains are buried beneath later sediments, but there is little doubt that all the massifs mentioned are parts of a belt which is continuous in depth.

The succession of older mountain systems in Asia south of Lake Baikal has been mentioned in



Fig. 7 Idealized cross section of the Appalachians from Ohio to the Atlantic

Plateau of Iran and including the Elburz Mountains at the south end of the Caspian Sea), the mountains of Afghanistan and western Pakistan to the Indus River, and the great Hindu Kush are included. This great belt stretches uninterrupted for 2500 miles, and in Iran is about 800 miles wide. It *is a continuation of the Alpine Cordillera and con*sists of mountains built chiefly during Tertiary time.

Himalayan Cordillera. The Himalayan Cordillera as here defined starts with the complex knot of mountains known as the Hindu Kush and the Pamirs on the west and extends eastward through the high and beautifully arcuate mountain chain of the Himalayas At the east end of the Himalayas the ranges veer sharply southward to pass through Indochina and down the Malayan Peninsula

The world's most extensive development of mountain systems is included between Lake Baikal in northeast central Asia and the Himalayas on the south The ranges that wrap around Lake Baikal were first formed during mid-Paleozoic time (Caledonian orogeny). The belt of compressional deformation extended as far south as the eastern Altai Mountains in Mongolia

Another belt of compressional mountains was added on the south in late Paleozoic time (Hercynian orogeny) This belt included the western Altai Mountains, the Tien Shan, and the bordering ranges of Tibet In mid-Mesozoic time the Tsinling Shan, which extends eastward to the Hwang-Yangtze delta plain, was built, and finally the great arcuate belt of the Himalayas was welded onto the growing continent in Tertiary time.

The Ganges plain on the south marks the site of an exceedingly deep basin which sank as the erosional waste from the Himalayas collected Tremendous earthquakes indicate continued crustal unrest and probably the continued rise of the great mountain mass. Although the geology of the Himalayas is only known fragmentarily, it is probably alpine in type.

East Indies arcuate complex. The islands of the East Indies from Sumatra to New Guinea extend over 3000 miles in an east-west direction and hence cover a region as large as the United States. They are divided geologically into four more or less concentric narrow arcuate belts of crustal deformation, each a separate fold and intrusive rock system. The centrally located and oldest, of Late Jurassic age, extends from Burma through Malaya into West Borneo. The next younger, of Cretaceous age, runs through Sumatra and Java and into southeast Borneo. The next, of mid-Miocene age, extends along the outer side of Sumatra and Java, around the inner arc of the Banda Sea, through the Celebes to Mindanao. The fourth is outside the mid-Miocene belt and runs from West Burma, through the Mentawai Islands, the Timor Ceram arc and the Celebes to Mindanao. It developed from Late Cretaceous to mid-Miocene time. The remarkable rows of volcanoes, the concentric submarine trenches, the belt of negative gravity anomalies, and the deep-seated earthquake foci all contribute to a world-wide interest in this region.

CIRCUM-PACIFIC CHAIN

The Circum-Pacific belt embraces the cordilleras of South and North America, the eastern Asiatic island archipelagos and the New Zealand-New Guinea systems.

North American Cordillera. The master division of the North American Cordillera is the Nevadan belt, which is typified by the Sierra Nevada of California Other divisions included in the North American Cordillera are the Rocky Mountains system, the Coast Ranges of California, and the mountain systems of Mexico.

Nevadan belt. The Sierra Nevada of California had its beginning in a thick accumulation of sediments and interlayered volcanic rocks. These were tightly folded and somewhat metamorphosed and then intruded by immense volumes of molten rock. When the molten material crystallized it formed great batholiths The term granite is commonly used to characterize the batholithic rock, but actually much of it contains less silica and potassium than does a true granite and should be called a granodiorite. The intrusions occurred from the beginning of Cretaceous to mid-Cretaceous time. The batholithic and metamorphic belt extends from Alaska to Tierra del Fuego with possibly only one break in Central America. *See* BATHOLITH.

Rocky Mountains system. This system lies inside (east of) the Nevadan (Fig. 6), and is characterized by folded and thrust-faulted strata of

taceous and Early Tertiary time. After considerable erosion of the folds and thrust sheets, the

The region of the Rocky Mountain belt between

at all angles The present ranges and valleys were largely blocked out by these faults and as a consequence, the Great Basin of internal drainage was

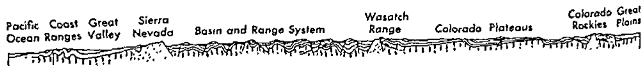


Fig. 6. Idealized cross section of Nevadan and Rocky Mountain belts from California to Colorado.

med Most of the major valleys are of structural rin rather than erosional.

untain belt welded ...

ces on the edge of the continent. They involve e western margin of the older Nevadan belt and ntain in places great thicknesses of Cretaceous d Cenozoic strata which accumulated after the tholithic intrusions. The San Andreas fault is e of a related system along which the oceanward ock of crust has moved horizontally northwestard for many miles. It is still very active

Mexico. The Nevadan belt extends southward e length of Baja California and probably rough the Sierra Madre del Sur The Gulf of California appears to be a downfaulted zone with the an Andreas fault system extending into it from e north. The Sierra Madre Occidental is a vast olcanic field over an underlying structural complex similar to the Rocky Mountains and Great Basin system of the United States The Sierra Madre Oriental is a continuation of Rocky Mountain structure into east-central Mexico from west Texas and New Mexico

South American Cordillera. This cordillera although supporting much higher peaks, is considerably narrower than the North American Cordillera in the United States Its *dominant element* is the Nevadan batholithic and metamorphic belt which borders the *Pacific coast for most of the distance* from Columbia to Tierra del Fuego Inland and flanking the *Nevadan belt is a fold and thrust* belt of younger age, somewhat like the Rocky Mountain *system but* generally narrower In places the two belts are separated by long narrow downdropped fault blocks, but in other places the faulting transgresses both belts Associated with the faulting, *if not* everywhere geographically then at the same time (Tertiary), are large outpourings of volcanic *rock* which extensively cover the batho-

mountain chain of late Mesozoic and Cenozoic age

of the Aleutians the belt is traced through Kamchatka and the Kuriles to the Japanese archipelago, and then southward in two major spurs, one along the Ryukyus (Nansei Shoto) to Formosa and the Philippines, *and thence to the* Dutch East Indies complex; the other through the Bonine and Mariana arcs and Yap and *Palau to Halmahera* of the East Indies This is a region of vigorous mountain growth at the present. Its characteristics are discussed under island arcs.

The wide belt of mountains occupying nearly all of Alaska, except the Arctic Coastal Plain, extends into Siberia. Recent U S.S.R. maps show that the belt curves southward around a small central shield and projects under the Okhotsk Sea. Evidently, *this main* and somewhat older cordillera lies submerged under the Okhotsk and Japanese Seas *between the mainland and the island archipelagos.*

OLDER CORDILLERAS

Major mountain systems older than those heretofore described occur extensively in North America, Eurasia, Australia, and Africa. The Appalachian Highlands include rather ancient systems in Newfoundland, the Maritime Provinces, and New England where the times of mountain building were Late Ordovician (the Taconic orogeny) and Late Devonian (the Acadian orogeny). The classical Appalachians of the eastern United States, whose parallel flat-topped ridges are erosional remnants

of the Taconic and Acadian belts of New England.

In Europe, most of Norway is a mountain belt of *Late Silurian age (the Caledonian orogeny) It* projects southwestward across the central and northern parts of the *British Isles. Discordantly* superimposed on the Caledonian belt and lying across southern England is a later belt of Pennsylvanian and Permian age, the Hercynian This spreads through most of France and parts of Germany north of the Alps and Juras.

The foundations of Brittany, the Vosges, the Ardennes, and the central plateau in France are Hercynian The same is true of the Black Forest, the Harz Mountains, and the Bohemian massif of Germany Most of Spain is Hercynian mountain structure and basement elements of the younger Pyrenees and Alps, such as the Mont Blanc and Aiguilles Rouges massifs, reveal Hercynian structures

The depressed regions between these various remnants of late Paleozoic mountains are buried beneath later sediments, but there is little doubt that all the massifs mentioned *are* parts of a belt which is continuous in depth.

The succession of older mountain systems in Asia south of Lake Baikal *has been* mentioned in

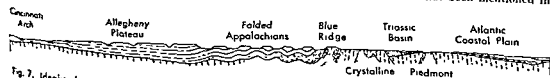


Fig. 7. Idealized cross section of the Appalachians from Ohio to the Atlantic

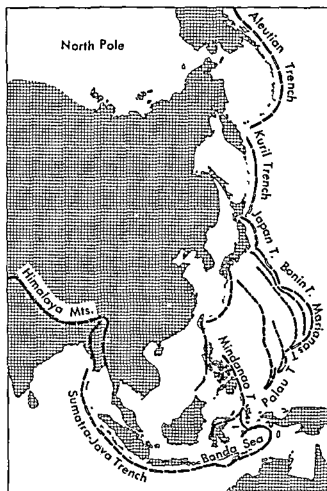


Fig 8 Island arcs and deep trenches of the western Pacific.

the discussion of the Himalayan Cordillera. The Urals of Russia are also an ancient mountain system of Paleozoic age.

In the sites of the late Mesozoic and Cenozoic cordilleras of North and South America the building of mountains during Paleozoic time is well recognized. In the Pacific marginal belts of the Americas mountain building has been going on from at least mid-Paleozoic time to the present. It reached climactic activity in the Nevadan orogeny of Early Cretaceous time.

ISLAND ARCS

Between some of the continents there are connecting mountain belts of the island-arc type, such as the Antilles between North and South America, the islands and submerged arc of the Scotia Sea between South America and Antarctica, and the Dutch East Indies between Asia and Australia. The smoothly curved festoons of islands in the western Pacific from Alaska to the East Indies have attracted great interest for many years, and now, with modern oceanographic and geophysical instrumentation, they are again an intriguing subject (Fig. 8).

Characteristics. In the early stages of growth an island arc is a row of volcanic cones built on a great curved swell of the ocean floor; the tops of the cones generally protrude above water and form the islands (*see* OCEANIC ISLANDS). The arcuate rows of volcanoes are convex toward the ocean, and on the convex side of the island arc is a parallel trench. Whereas the general ocean floor is 12,000–15,000 ft deep, the trenches reach depths of 25,000–35,000 ft (*see* SUBMARINE TOPOGRAPHY). In places there are two subparallel arcuate rows of islands, one back of the other, and in the submarine topography even a third swell may be suggested. These are regions not only of many active volcanoes, but of repeated earthquakes, so that there is little doubt that the crust is being energetically deformed. The sites of the trenches have been found to be the positions of a great deficiency in gravity, and from this observation it is deduced that the light-weight upper crustal layers have been downfolded into the underlying heavier subcrustal and mantle rocks. As long as forces, whatever they are, hold down the thickened crust, there will be a negative anomaly of gravity as measured at the surface. As soon as the forces relax, the downfold will adjust upward like a buoyant ship rising in water as its cargo is unloaded. A mountain range may appear out of the ocean in the place where formerly there had been a trench and a volcanic island arc. *See* TERRESTRIAL GRAVITATION.

The positions of shallow, intermediate-depth, and deep-seated earthquakes have been charted in relation to the island arcs and trenches. These are found to be approximately along a great curved surface which dips downward at about 45° under the continent from the trench to depths of nearly 700 km (Fig 9). The integration of the geological and geophysical data has not been satisfactorily accomplished as yet although a number of attempts have been made. Nevertheless, it is believed that the continental cordilleras will be better understood when the island arc mysteries are solved

Sial and sima. These terms are used to depict the upper layer and the lower layer, respectively, of the crust. The sial is composed of all the rocks exposed at the surface and is high in silica and the alkalic elements The sima is leaner in silica and richer in iron and magnesium The sial has a density of about 2.7, whereas the sima is about 3 0. Below these layers is the great shell called the mantle. It has still less silica and more iron and magnesium than the sima, its density is about 3 3 immediately below the sima and increases with depth. Figure 10 illustrates the constitution of the crust of the continent and the ocean basin in idealized form Seismic waves are reflected and refracted from the boundaries of the layers The depth of these discontinuities has been shown to vary; it is generally deeper under major mountain ranges and shallower under the plains and stable regions.

Andesite line. This line has been drawn to separate the regions of andesitic volcanic rock from those of basaltic volcanic rock in the western and southwestern Pacific. It suggests what is probably a significant concept. Andesite is a volcanic rock of higher silica content than its neighbor and common associate, basalt. Basalt is believed to come directly from the subcrust or sima, which is also

ered to be basaltic in composition. Andesite, he other hand, comes from a molten rock which ither a basalt contaminated with material from sial or has been produced entirely by the melt- of the lower part of the sial. Andesite, there- e, is believed to mark continental conditions and haps, more specifically, mountain belts whose ots are melting. Basalt is commonplace in andes- terranes, and may be emitted intermittently th andesite from the same volcano, such a locale considered to be that of a continental mountain lt. In contrast, if only basalt is erupted, the locale considered to be oceanic.

The andesite line separates the andesitic volca- oes of the western Pacific island arcs from those f the basaltic Hawaiian Islands, and from the ba- altic Midway, Wake, Marshall, Gilbert, Ellice, amoa, and the Caroline islands Within the an- esite province are the Bismark, Solomon, New Hebrides, New Caledonia, Fiji, Tonga, Kermadec, Chatham, and Auckland islands, as well as New Zealand

SHIELDS AND STABLE AREAS

Shields and stable areas are those parts of continents which for considerable geologic time have remained fairly undisturbed and have not been involved in mountain building The interior and northern parts of North America between the Rocky Mountains and the Appalachians is known as the stable interior or the "craton." Part of the stable interior consists of ancient Precambrian rocks at the surface, and is called a shield, such as the Canadian shield. Part of the stable interior consists of a basement of Precambrian rocks blanketed with a veneer of sedimentary rocks. Within this veneered part there are many broad basins and domes. *See* PRECAMBRIAN.

Some geologists have emphasized a theory that

ble for the continental margins that border the Pacific, but aside from North America with its Appalachians, the continents that border the Atlantic and Indian oceans do not have marginal belts In a number of places, mountain belts like the Caledonian of Great Britain and the Hercynian of Brittany, trend normal to the coast line and disappear under the ocean

Rift valleys are long narrow depressions, generally in stable or shield areas, which have been formed by relative down-dropping of a wedge-shaped block of the earth's crust. It is possible that the adjacent crustal blocks rose more than the wedge block sank The fault valleys formed are also called graben. Thus we speak of the Rhine graben, 30 miles wide and 200 miles long The classical rift valleys are those in Asia Minor and East Africa. They include the depressions occupied by the Dead Sea, the Red Sea (including the Gulfs of Suez and Aqaba), and the deep, lake depressions of Tanganyika and Nyasa. The great East African rift valleys extend nearly 2000 miles from south to north, and the faults cut through all types of rocks. Some volcanism is present, seemingly connected with the faulting. The origin of the rift valleys is still unsettled. *See* RIFT VALLEY; *see also* CORDILLERAN BELT, MOUNTAIN SYSTEMS. [A J.E.]

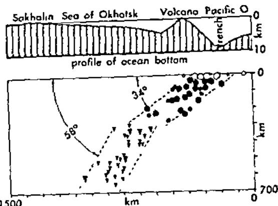


Fig 9. Earthquakes of the Kamchatka-Kurile region projected to an intermediate vertical plane (After H Benioff, Geol Soc. Am. Spec. Paper 62, 1955)

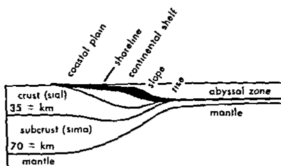


Fig 10 Sial and sima at the continental margin. Black denotes unconsolidated sediments, stippled pattern denotes semiconsolidated material.

Bibliography: L. U. De Sitter, *Structural Geology*, 1956; A Holmes, *Principles of Physical Geology*, 1945.

Tectonophysics

The science of the physical processes involved in forming geological structures. It is part of an older branch of geology called tectonics.

The application of physics to geological problems has enabled tectonophysics to provide a deeper understanding of the earth in three ways: description is now possible of the divisions of the earth's crust and the substrata of the upper mantle as well as of the land surface; some theories of the nature and rates of processes within the earth have been proposed, and the location of tectonic forces has been discovered to lie at depths of tens or hundreds of kilometers rather than at the surface.

DESCRIPTION: UPPER MANTLE AND CRUST

The chief divisions of the interior of the earth, determined by studies of the elastic waves gener-

ated by large earthquakes, are the solid crust, the solid mantle, and the liquid core. *See* EARTH INTERIOR; SEISMOLOGY.

Crust. The base of the crust lies at a depth of from 5 to 60 km. The crustal envelope may be divided in two stable parts, the ocean floors and the continents, and two active parts, the mid-ocean and continental fracture systems.

Mantle. According to K. E. Bullen, this internal zone is concentrically layered with chemical, not phase, changes at depths of 413 and 984 km. It probably consists of impure magnesium and iron silicates containing about 40% SiO_2 and known as ultrabasic rocks. Unfortunately, the abundance of heat-producing, radioactive elements, which is particularly important in tectonophysics, is unknown.

Core. There is no evidence that the core has had much influence upon the earth's surface features, which are portrayed in the crust and are considered to be molded by actions in the outer part of the mantle.

Ocean floors. Covered by an average of 5 km of sea water, ocean floors form more than half the earth's crust. They appear to be structurally simple, uniform in character, and probably very old. A typical section consists of 1 km of sediment, having a seismic-wave transmission velocity of about 2 km/sec and a density of 2.3 overlying a layer 4.5 km thick of basalt. This basalt has a velocity of about 6.5 km/sec, a density of 2.85, and a content of about 50% silica.

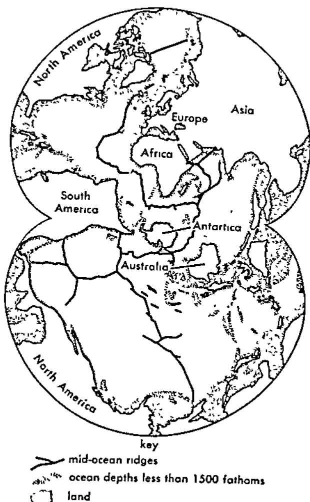


Fig. 1. Mid-ocean fracture system and ridges.

Mid-ocean fracture system. This system follows ridges upon the ocean floors, beneath which lies the second most active seismic system on earth. In 1956, W. M. Ewing and B. C. Heezen combined bathymetric and seismic evidence to suggest that the mid-ocean ridges form a continuous system about the earth.

Where it has been examined, the ridge feature consists of a broad swell several hundred kilometers wide, with a rugged crest rising from 3000 to nearly 11,000 m above the deep ocean floors. Together with its known branches, this system is over 40,000 km long, and the main part everywhere maintains a position as far removed as possible from continental margins. Where the ridge has been closely studied, a central longitudinal rift has generally been found, beneath which lie earthquake foci whose depths never exceed 70 km. Beneath discontinuous shallow pockets of sediment, the upper most 3 to 5 km of the ridge has an average seismic-wave velocity of 5.2 km/sec, identified as basalt. Beneath this, a deeper layer, perhaps a mixture of basalt and ultrabasic rocks, with an average velocity of 7.2 km/sec, forms a root which may extend to 30 km below sea level. Gravity measurements suggest that the ridges are in isostatic equilibrium.

The ridges appear to have been accumulating in their present positions for a long time, perhaps as long as the continents.

Continental fracture system. The most active part of the earth's surface, this system comprises the seismically active mountains and island arcs at the borders of continents. It has a T shape and lies around much of the margin of the Pacific Ocean and crosses southern Eurasia.

Analysis of the geological and geophysical features of this system has suggested that it is made up of two types of elements called primary and secondary arcs, each repeated many times in various stages of evolution.

Primary arcs. Such arcs are volcanic and igneous mountain chains or island chains, of which a typical cross section is illustrated. They have these distinguishing characteristics.

1. Their shape is circular, concave toward the nearest continent.

2. They have recent andesitic volcanic and older granodioritic igneous activity (rocks with about 60% SiO_2).

3. All the deepest trenches in the oceans parallel them.

4. Large negative gravity anomalies occur along them in narrow strips (*see* OROGENY; TERRESTRIAL GRAVITATION).

5. Most of the world's shallow earthquakes and all the world's deep earthquakes (from 70 to 700 km deep) occur beneath them.

6. They rest upon no basement of more ancient rocks.

The first four of the primary arcs listed in Table 1 are regular and are believed to be stages in an

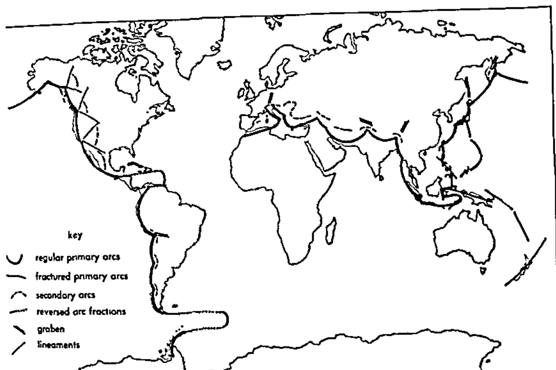


Fig 2 The continental fracture system of active mountains and island arcs.

evolutionary sequence by which small island arcs grow through volcanism, metamorphism, and up-

less igneous, and less metamorphosed They are convex toward the continent and consist of a welt of uplifted older basement rocks, beyond which is a basin of sedimentary rocks, folded and thrust on to the continent. Although they form some of the great mountains of the world, they lack the geophysical evidence of deep connections which distinguish primary arcs

Arc junctions. Such junction patterns are of three main types—reversed arc, single graben, and double lineament.

Reversed arc junction. The simplest is the reversed arc

systems in this, the junction of two primary arcs is capped by a secondary arc beyond which one graben or broad structural valley radiates from the junction. Examples occur at the Alps and in Bolivia, where the Rhine and Chiquitos grabens are well known *See* GRABEN.

Double lineament junction. The third type of junction is the double lineament junction, giving rise to broad mountain systems. In this case, two lineaments radiate from the junction of two primary arcs and enclose the secondary arc between them at a distance of several hundred kilometers from the junction. A lineament is a disturbed zone along which there is faulting and across which changes occur in rock facies and structure. The Cordillera of North America provide the best examples.

Related continental patterns. The continents cover more than one-quarter of the earth's surface and have a thickness of from 30 to 60 km. Continental crustal thickness is determined both by the refraction of seismic waves from large explosions and by measurement of changes in phase velocities of surface earthquake waves. They are blocks of gneissic rock with a composition of 60-70% of SiO_2 overlain by well-sorted sedimentary rocks which, only in narrow troughs or geosynclines

Table 1. Active primary arcs and their features

Name	Sedimentary part	Igneous part	Example
Single island arc	Trench	Volcanic islands	Kuril Is.
Double island arc	Sedimentary	Volcanic islands	Aleutian Is at Kodiak I.
Single mountain arc	Trench	Igneous ranges	Central Andes
Double mountain arc	Sedimentary	Igneous ranges	Coast, Cascade, and Sierra Nevada Mtns
Fractured straight island chain	Irregular features	Irregular features	Solomon Is.

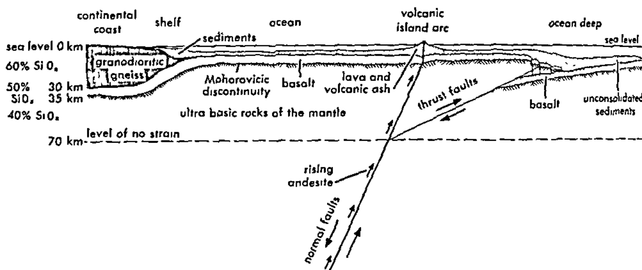


Fig. 3. Cross section of single island arc.

reach a thickness of more than a few hundreds or thousands of meters.

Until recently, it seemed natural that continents should be regarded as permanent, but consideration of the rates at which geological processes act now indicates that they have grown. In 1927, K. Sapper pointed out that volcanism was resulting in extrusion of lava, mainly andesite, at a rate of at least 0.8 km^3/yr. This was thought to be due to the downwarping and remelting of continental margins along the continental fracture system, until the discovery that such marginal seas as Bering, Caribbean, or Tasman seas are underlain by oceanic crust showed that the andesites were new accessions of crustal material. The rate is approximately sufficient to have extruded all the crust in geological time. The suggestion that this has happened is supported by the proposals by W. W. Rubey and others that all the oceans and atmosphere are also products of volcanism.

Recent measurements of the rates of formation of sediments and comparisons between the relative volumes of recent and ancient sediments suggests that lavas are eroded to sedimentary rocks, which accumulate along continental margins where they are ultimately metamorphosed to gneissic rocks during the formation and evolution of primary arcs On other margins of continents, older, inactive mountains like the Appalachians, display similar patterns to existing active mountains This suggests that the continental fracture system occasionally moves from one margin to another, thus enlarging the continents first on one side, then on another.

The loads thus formed upon the surface, slowly settle into approximate hydrostatic balance, but after extrusion, the rocks assume less dense forms, and thus the continents stand high (*see* ISOSTASY).

The development of radioactive methods of age determination has radically lengthened and changed our conception of Precambrian time. It is now apparent that continents are not primeval blocks, but that each is zoned and that each zone may be a former location of part of the continental fracture system.

Thus, the fundamental problems in tectonophysics are to provide theories for the location, formation, and structure of the mid-ocean and continental fracture systems.

MECHANICAL BEHAVIOR OF EARTH MATERIALS

Before considering the forces in the earth's crust and the major structures of the crust, it will be well to consider the rate at which deformation proceeds and the mechanism of those smaller deformations, faults and folds, which have shaped the crust.

Dynamics of faulting. Rock fractures are of two classes: joints or tension cracks, which are partings without appreciable relative movement of the two sides, and faults, which are shear failures. Faults have been divided into three classes whose properties are given in Table 2

Faults are an expression of localized mechanical failure of the material of the earth's crust. The classical theory due to O. Mohr has been used by

Table 2. Characteristics of the three classes of fault

Property	Normal	Transcurrent	Thrust
Vertical axis	Major	Intermediate	Minor
Dip	65° and straight	90° variable	25°
Strike	Wavy	Straight	Very wavy
Direction of movement	Down dip	Horizontal	Up dip
Nature of fracture	Brecciated	Indeterminate	Sheared
Intrusives	Dikes possible	Uncommon	None
Connections with mountains	None; gravity greatest force	Yes; faults at 25° to forces	Yes; faults normal to forces

d Anderson to explain the different types of s.

er a broad area, the surface of the ground can onsidered to be plane. Since it is a surface of [illegible] of stress [illegible], and [illegible] P [illegible] Q [illegible] R, where sses are considered to be positive if pressures, tive if tensions Three possible relations then t which correspond to the three types of fault explain their characteristics.

ormal faults The greatest pressure, P, is verti- In general, the horizontal stresses will not be al, and if failure occurs, it will take place along lane parallel to Q, inclined at an angle $\phi < 45°$ h the vertical. Thus, the fault planes dip at gles greater than 45°, striking at right angles he direction in which relief of pressure is great- The motion is such that the horizontal extent ncreased and is characteristic of a normal fault.

Transcurrent faults. If the intermediate prin- al stress Q is vertical and failure occurs, it st happen in a vertical plane inclined at an gle $\phi < 45°$ with the greatest pressure P. The Q is vertical, and movement is nearly horizontal, aracteristic of a transcurrent fault.

Thrust faults. If the minimum pressure R is rtical when failure occurs, it will take place ong a plane parallel to Q, inclined at an angle $< 45°$ with the direction of P Thus, the fault anes will have shallow dip, and motion such that e horizontal extent will be shortened. In this case, e characteristics of a thrust fault are explained.

Dikes Although dikes are found along some ormal faults, this is not usual Dikes are much less ommon than faults, and most follow tension cracks hich form parallel to planes of no shear *See* ULTON.

Dynamics of folding. The folding of sedimen- ary beds and the formation of mountain ranges show that crustal rocks may be highly deformed in a manner which could result only from plastic flow (*see* ROCK MECHANICS).

Under the ordinarily familiar conditions and lengths of time, most rocks behave as very brittle solids. Therefore, the plastic deformation of rocks must take place over greater lengths of time and at higher temperatures and pressures While the temperatures and pressures involved in near-surface phenomena can be reproduced in the laboratory, the enormous intervals of time involved preclude the possibility of directly observing the plastic deformation of rocks in geology Nevertheless, many model experiments have been carried out. Where only mechanical phenomena are to be considered, the scale of a model is related to the original by three arbitrarily chosen parameters. Mass, length, and time are usually chosen, and the ratio of all other physical properties can be expressed in terms of the scale ratios m, l, and t In problems of geologic nature, forces due to gravity are of prime importance, and the acceleration due to gravity is usually the same in the model as in the original. Thus, the product $lt^{-2} = 1$, or, $l = t^2$.

In consequence, if the time ratio is suitably chosen, the length ratio usually renders the model microscopic in size. Conversely, the time scale imposed by a convenient choice of linear dimensions is usually far too great. Moreover, in many cases, correct scaling ratios require the model to be made of materials having unrealizable physical properties.

Although most geologic problems cannot be scaled down to laboratory experiments, the results of such experiments as can be performed lend general support to the supposition that the apparently brittle and rigid nature of surface rocks is not inconsistent with rock flow over great distances and over long intervals of time to give rise to folded rocks and mountains.

Important experimental work with models has recently been described by V. V. Beloussov, W. H. Bucher, D T Griggs, M V. Gzovski, E Cloos, and J W. Hardin.

Rates of tectonic deformation. A. E. Scheidegger has classified stresses according to whether their duration is short, intermediate, or long The upper limit of stresses of short duration is about 4 hours In this range, the material of the crust and mantle behaves as an elastic solid with a rigidity of about 2×10^{12} dynes/cm² and a Young's modulus of about 5×10^{12} dynes/cm² If the elastic limit is exceeded, the material undergoes brittle fracture This information is obtained from laboratory experiments and the passage of earthquake waves

The time range of stresses of intermediate duration is from about 4 hours to 15,000 years. Information on this time range comes from two sources: the damping of the earth's free nutation and the release of stress in earthquake sequences If it is assumed that the mantle behaves as a Kelvin body, exhibiting elastic aftereffect, it follows that the

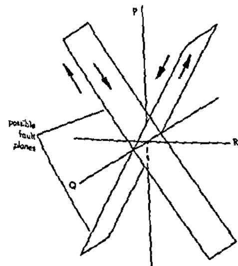


Fig. 4. Orientation of stress axes and normal faults.

two Kelvin constants corresponding to rigidity and viscosity are

$$\mu_K = 2 \times 10^{12} \text{ dynes/cm}^2$$

and

$$\eta_K = 3 \times 10^{17} \text{ g/sec-cm}$$

Kelvin, Maxwell, and Bingham bodies are theoretical models of materials following particular laws of rheological behavior (*see* ELASTICITY; RHEOLOGY). The occurrence of earthquakes shows that the Kelvin type behavior exists only up to a certain limiting stress. If that limit is exceeded, fracture occurs. It is not possible to say what the limiting stress is, or what the mode of fracture is. There is no justification for assuming that the results of experiments of short duration with rocks in the laboratory will apply to stresses in the intermediate range.

Finally, there is the problem of the rheological behavior of the material of the earth's crust and mantle under stresses of long duration, up to hundreds of millions of years In this range, creep is the dominant characteristic, and even less is known about behavior At the lower end of the time interval is the rise of some land masses due to isotasy since the melting of the ice at the end of the last ice age (*see* WARPING, EARTH CRUST) Estimates of the postglacial uplift of Fennoscandia are of the order of centimeters per century, corresponding to a value of η_M of about 10^{22} gm/cm-sec, where η_M is the Maxwell body constant corresponding to viscosity. Thus, the earth's mantle and crust show *creep effects with a relaxation time of the order of* 20,000 years. However, mountain ranges can persist over several million years, their eventual disappearance being the result of erosion and not of creep. Hence, the material must also exhibit a yield stress, so that it approximates a Bingham body more closely Estimates of the yield stress for the surface of the earth give 4×10^9 dynes/cm²; at a depth, it may well be lower. In general, the creation of an orogenetic system will take place over millions of years, and it is thus incorrect to use such attributes as rigidities and viscosities calculated from short and intermediate data to explain such phenomena On the other hand the laws of the intermediate time range govern local tectonics, such as folding and faulting, and those of short duration govern the passage of earthquake waves. It is this varying response of the earth to stresses applied over different time intervals which makes the study of tectonophysics difficult and *which has caused much confusion in the past*

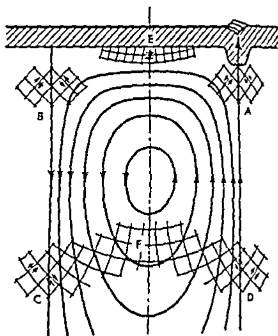


Fig. 5. Vertical cross section of convection current showing areas of maximum shear stress beneath a primary arc. (After F. A Vening Meinesz)

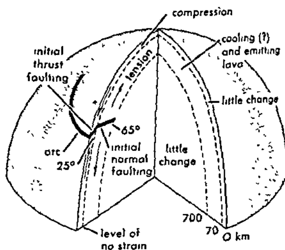


Fig 6 Arc formation in a contracting earth

Rock failure. The forces and patterns of geological failure in the earth's crust and upper mantle will be considered together There is as yet no agreement upon which of many possibilities is the cause of the earth's surface structure.

Polar wandering. The hypothesis of polar wandering depends upon the nearly spherical shape of the earth and upon its supposed ability to yield to long-term stress. If, from time to time, large masses such as uplifted mountains or ice sheets are placed eccentrically upon the earth's surface between the poles and the equator, they will produce forces tending to make the entire globe or the crust and upper mantle move as a unit relative to the polar axis. It seems possible that this has happened, and it may explain some paleoclimatic and paleomagnetic observations, but no tectonic effects have yet been demonstrated. F A Vening Meinesz has endeavored to show that many structural fractures fit the shear pattern which would be caused by a movement of the poles over 70° along the meridians of 90° longitude Unfortunately the features which do fit the pattern are of very diverse ages, and few of them are due to shearing There are large gaps in the pattern unrepresented by any features, and other equally important features are not explained

Continental drift This theory, now largely superseded, assumes a drift between parts of the crust. This assumption arose from evidence of past climatic changes, the supposed need for migration routes for ancient animals and plants, the great

shortening observed in surface folds in Alpine mountains, parallelism of the coasts of the Atlantic Ocean, alleged similarities in stratigraphy between opposite coasts, and recent measurements of paleomagnetism. The evidence for the last two is disputed, and the other reasons may have other explanations. *See* SUBMARINE TOPOGRAPHY.

Thermal convection currents. The theory that the major structures of the earth's crust are due to subterranean thermal convection currents has recently been elaborated by W. A. Heiskanen and F. A. Vening Meinesz. *See* EARTH (HEAT FLOW). They believe that intermittent currents with changing patterns have circulated throughout geological time below a depth of a few tens of kilometers. They consider that early in the earth's history, a first order convection system brought about the formation of the core and of a primitive sialic crust in the form of one continent.

During the next stage, second and third order convection systems occurred through the whole mantle and rose under the present oceans to separate the continents. Later, higher order systems acted intermittently in the upper part of the mantle to form smaller tectonic features.

Convection current theories have been popular because they can explain the loss of heat by the mantle (in which radioactivity and temperatures are unknown), and the negative gravity anomalies along island arcs (now considered to be due to light sediments), and because they can be treated mathematically. Difficulties arise, however, in trying to explain deep-focus earthquakes which arise from stress differences in terms of flow.

Correlation with geology has always been in general terms and many contradictions are unresolved. Thus, the Mid-Atlantic Ridge is attributed to rising currents and tension by some authorities, but to sinking currents and compression by others. In either case, the pattern of currents is hydrodynamically difficult to explain at the bifurcations of the ridge in the Indian Ocean (see Fig. 1). Those favoring compression can find support in Pakistan where one branch of the ridge enters Pakistan as a thrust-faulted range, while others would emphasize that another branch joins the African rift valleys which they consider to be due to tension (*see* RIFT VALLEY).

Contraction theory. Another important theory is the contraction theory, which once depended upon cooling in the mantle, but which can now be seen to be a necessary corollary of the growth of continents. If the Mohorovičić discontinuity represents the original surface of the earth upon which has been extruded the crust and oceans with an average thickness of 20 km, the circumference of the original surface must have shrunk by 125 km.

This contraction is considered to be the cause of the two fracture systems. No very striking characteristics have been observed in the mid-ocean system, but the continental system has one outstanding peculiarity: it is formed of a scalloped series of conical fractures. Thus, the basic problem of mountain building is how a conical fracture can be formed on the surface of a shrinking sphere.

The innermost zone, extending from the center to within 700 km of the surface, is, he thought, not changing in temperature nor volume. The next zone, extending from 700 to 100 km, is cooling most actively and hence contracting and stretching about the inner one. The outer zone above 100 km has already largely cooled, and it is losing support and hence is becoming compressed by the contraction of the intermediate zone beneath.

This view is not altered in its essential points if the activity of the earth is regarded as being due primarily to emission of lava by, and hence shrinking of, the same intermediate zone. Nor is it altered if we use the levels at which earthquakes cease

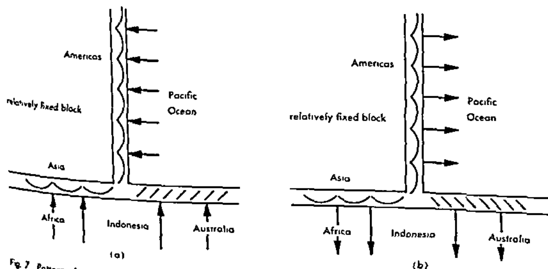


Fig. 7 Pattern of continental fracture system resulting from (a) compression and (b)

two Kelvin constants corresponding to rigidity and viscosity are

$$\mu_K = 2 \times 10^{12} \text{ dynes/cm}^2$$

and

$$\eta_K = 3 \times 10^{17} \text{ g/sec-cm}$$

Kelvin, Maxwell, and Bingham bodies are theoretical models of materials following particular laws of rheological behavior (*see* ELASTICITY; RHEOLOGY). The occurrence of earthquakes shows that the Kelvin-type behavior exists only up to a certain limiting stress. If that limit is exceeded fracture occurs. It is not possible to say what the limiting stress is, or what the mode of fracture is. There is no justification for assuming that the results of experiments of short duration with rocks in the laboratory will apply to stresses in the intermediate range.

Finally, there is the problem of the rheological behavior of the material of the earth's crust and mantle under stresses of long duration, up to hundreds of millions of years. In this range, creep is the dominant characteristic, and even less is known about behavior. At the lower end of the time interval is the rise of some land masses due to isotasy since the melting of the ice at the end of the last ice age (*see* WARPING, EARTH CRUST). Estimates of the postglacial uplift of Fennoscandia are of the order of centimeters per century, corresponding to a value of η_M of about 10^{22} gm/cm-sec, where η_M is the Maxwell body constant corresponding to viscosity Thus, the earth's mantle and crust show creep effects with a relaxation time of the order of 20,000 years However mountain ranges can persist over several million years their eventual disappearance being the result of erosion and not of creep Hence, the material must also exhibit a yield stress, so that it approximates a Bingham body more closely. Estimates of the yield stress for the surface of the earth give 4×10^8 dynes/cm², at a depth, it may well be lower In general the creation of an orogenetic system will take place over millions of years, and it is thus incorrect to use such attributes as rigidities and viscosities calculated from short and intermediate data to explain such phenomena On the other hand the laws of the intermediate time range govern local tectonics, such as folding and faulting and those of short duration govern the passage of earthquake waves It is this varying response of the earth to stresses applied over different time intervals which makes the study of tectonophysics difficult and which has caused much confusion in the past

Rock failure. The forces and patterns of geological failure in the earth's crust and upper mantle will be considered together. There is as yet no agreement upon which of many possibilities is the cause of the earth's surface structure

Polar wandering. The hypothesis of polar wandering depends upon the nearly spherical shape of the earth and upon its supposed ability to yield to long-term stress. If, from time to time, large masses such as uplifted mountains or ice sheets are placed eccentrically upon the earth's surface between the poles and the equator, they will produce forces tending to make the entire globe or the crust and upper mantle move as a unit relative to the polar axis It seems possible that this has happened, and it may explain some paleoclimatic and paleomagnetic observations, but no tectonic effects have yet been demonstrated F A Vening Meinesz has endeavored to show that many structural fractures do fit the pattern are of very diverse ages, and few of them are due to shearing There are large gaps in the pattern, unrepresented by any features, and other equally important features are not explained.

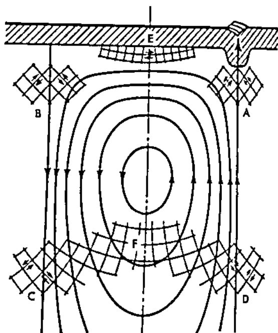


Fig 5. Vertical cross section of convection current showing areas of maximum shear stress beneath a primary arc. (*After F A Vening Meinesz*)

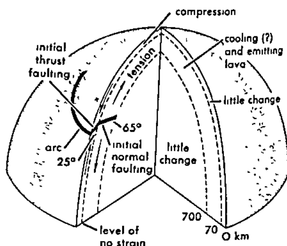


Fig 6 Arc formation in a contracting earth

Continental drift This theory now largely superseded, assumes a drift between parts of the crust. This assumption arose from evidence of past climatic changes, the supposed need for migration routes for ancient animals and plants, the great

…mical composition of tektites, obsidian, … Libyan Desert glass

…ound	Range in composition for tektites, %	Average for 73 tektites, %	Range in composition for obsidian, %	Average for 32 obsidians, %	Libyan Desert glass, %
[illegible]	68 00–89 81	74 64	66 70–78 15	70 40	98.20
[illegible]	6 13–15 42	12 12	11 54–18 30	15.26	0 70
[illegible]	0 00–2.25	0 60	0 09–3 92	1 55	0 53
[illegible]	0.89–6 46	4.20	0 08–2 87	0 57	0 24
[illegible]	0 22–4 05	2 03	Trace–1 08	0 29	0 01
[illegible]	0 04–3 92	2 45	0 28–2 67	1 11	0 30
[illegible]	0 10–2.46	1 22	0 67–6 50	4 02	0.33
[illegible]	0 82–3 60	2.23	1 45–5 38	4.24	0 02
[illegible]	0 18–1 40	0 82	*	*	0 33
[illegible]	0 001–0 01	*	0 8–1 0	0.2	0 06

* Data not acceptable for averaging

…e Philippine Islands, and in the lands adjacent …the South China Sea. They also have been found …n Bohemia; the Ivory Coast of Africa, Colombia, …nd in Georgia and Texas Tektites are associated …ith Eocene, Miocene, and Recent formations, in…licating they either fell when these beds were ac…cumulating or were made within them by impacts of …meteorites.

Origin. The form and distribution of tektites suggest that small masses of viscous material entered our atmosphere, or formed therein before falling to the earth How or where tektites originated is a mystery Theories proposed are (1) tektites represent siliceous material melted from stony meteorites passing through the earth's atmosphere; (2) tektites are fused material thrown from the moon after either a meteorite or a comet collided with that body; (3) they were formed by a meteorite striking the earth; (4) they are fragments of an exploded planet; (5) tektites are the debris of a comet which passed close to the sun, (6) they were formed by the fusion of earth material when a comet collided with the earth's atmosphere

If tektites are extraterrestrial in origin, several …

… meteorites, they apparently were not formed from meteoritic matter. This observation, together with the fact that the potassium-argon age of tektites is the same as the sediments on which they are found, challenges the extraterrestrial origin of tektites See METEORITE; ROCK (AGE DETERMINATION)

[E.P H.]

Telecast

A television broadcast, involving the transmission of the picture and sound portions of the program …

… reception by the general public, just as is a radio broadcast. The picture may be either in black and white or in full natural color, using amplitude modulation, while the sound portion (in the United States) uses frequency modulation. The channels assigned for telecasts by the Federal Communications Commission, each 6 megacycles (Mc) wide, cover frequencies as follows: 54–72 Mc (channels 2 through 4), 76–88 Mc (channels 5 and 6), 174–216 Mc (channels 7 through 13), and 470–890 Mc (channels 14 through 83). See TELEVISION.

[J.MR]

Telegraphy

A method of communication employing electrical signaling impulses produced and received manually or by machines. Telegraph signals are transmitted over open wire or cable land lines, submarine cables, or radio. Telegraphy as a communication technique uses essentially a narrow frequency band and a transmission rate adapted to machine operations. See COMMUNICATIONS, ELECTRICAL; DATA TRANSMISSION.

Early equipment devised by Samuel F. B. Morse consisted of a mechanical transmitter and receiver or register. Operators soon learned to handle messages faster using simple manual keys and audible sounders. Subsequently, telegraph transmission and reception again became mechanized. See TELETYPEWRITER EXCHANGE (TWX) SERVICE; TELEX. Telegraphy may also be used in other ways. See FACSIMILE; TELEPHOTOGRAPHY; TELETYPESETTER.

Telegraph facilities for use by the general public to transmit messages both domestically and internationally are provided by communication companies and government administrations. Special telegraph facilities include those for news services, distribution of market prices of securities and commodities, and private lines betweens such points as the factories and offices of a company for the exchange of messages, orders, payroll data, and warehouse inventories. Municipal and private fire and police alarms are a special form of telegraphy. The armed forces have extensive fixed and mobile telegraph systems.

Telegraph codes. For manual operation, the the code consists of short dot and long dash signals (Fig. 1) The original Morse code also used various length spaces; the Continental code avoids them. On submarine cables, the dots and dashes are of equal length for most efficient use of the transmission characteristics and are distinguished by being of opposite electrical polarity.

Most automatic printing telegraph circuits, including American cable operation, use a code of five equally spaced signals or units per letter or other symbol perforated into a paper tape (Fig. 2). The presence or absence of current or current reversals during these intervals constitutes the distinguishing feature. When a perforated tape is used, it usually is driven by a sprocket running in holes between the second and third code unit. Machines translate automatic teleprinter code to cable code, cable code to automatic code or, for special applications, make other translations. For stock quotation systems and teletypesetter opera-

(700 km) and at which a marked change in their nature occurs (70 km) as the boundaries of the zones. Cooling may or may not have occurred.

A. Ritsema has deduced from his studies of earthquakes that the stresses in the earth are equal in all horizontal directions. A. E. Scheidegger has shown that under these circumstances, according to O. Mohr's theory, conical fracture can occur, and that these cones should dip at about 20–30° near the surface if caused by compression and dip at about 60–70° at greater depths if caused by reduction of pressure. The location of earthquake foci is not very precise, but the evidence does suggest that this is the shape of conical fracture zones below island arcs.

The theory also explains fractured arcs, each of which has along it evidence of a great transcurrent fault like that along the axis of New Zealand

All the fractured arcs lie along one limb of the T of the continental fracture system It is not likely that this distribution is due to chance. The proposed explanation is that only that arm had to undergo two sets of movements approximately at right angles to each other. Along two of the three limbs of the T, contraction or compression can take place without horizontal shearing, but along the third arm, shearing must accompany movement

The direction of the fractured arcs will depend upon whether they are primarily formed by compression or by relief of pressure, and a comparison between Fig. 7 and Fig 2 shows that the cause of mountain building could have been relief of pressure in the zone between 70 and 700 km, as postulated.

The contraction theory can also explain the properties of broad and narrow mountain systems. Consider two primary arc fractures meeting in th zone of contraction; both are normal faults an tend to move apart, but cannot do so unless at leas one additional fracture occurs at the junction.

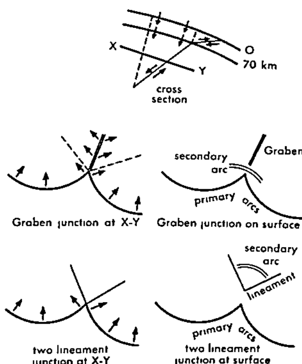


Fig. 8. Cross section and plans at depth and surface of the two types of arc junctions.

If one such fracture forms, it will tend to open allowing the rocks above to drop, forming a graben on the surface and the narrow type of mountain belt.

If instead, two shears form and radiate out from the junction, they appear as lineaments, and this forms the wide pattern of mountain belts.

Thus, the contraction theory, although not yet universally accepted, provides a physical explanation for far more tectonic details than any other theory. [J T W]

Bibliography: E. M. Anderson, *The Dynamics of Faulting*, 2d ed., 1951; W. A. Heiskanen and F. A. Vening Meinesz, *The Earth and Its Gravity Field*, 1958; J. A. Jacobs, R. D. Russell, and J. T. Wilson, *Physics and Geology*, 1959; A. E. Scheidegger, *Principles of Geodynamics*, 1958.

Tektite

A general term applied to irregularly rounded, comparatively small, glassy objects believed by some to have fallen through the earth's atmosphere from outer space. They are named according to the locality in which they are found; for example, tektites from Moldau River, Bohemia, are called moldavites; from Billiton Island, billitonites; from Australia, australites; and those from Grimes County, Texas, bediasites after a town, Bedias.

Physical form and properties. Tektites have a variety of shapes, although most of them are spheroids. Their surfaces may have a variety of irregularities such as grooves, conchoidal depressions, concentric equatorial rings, and pits. The arrangement of these features may suggest aerodynamic shaping.

Tektites usually are dark in color; however, the Bohemia and Georgia samples are light green. All tektites are brittle and have a conchoidal fracture. Their index of refraction is between 1.48 and 1.53, their specific gravity between 2.32 and 2.52.

Many tektites, if examined by transmitted light, show irregular colored zones and small inclusions of lechatelierite (fused quartz). Tektitic glass frequently shows internal strains and occasionally contains bubbles but is never scoriaceous. Obsidian differs from tektites by the types and arrangements of inclusions

Composition. The composition of tektites, ob-

are not chemically similar to the rocks with which they are found.

Occurrence. Tektites are more difficult to find than stony meteorites, and therefore, their distribution is not as well known as the distribution of stony meteorites. Tektites are numerous in Australia, with the greatest concentration in the south and decreasing northwards, in all of Tasmania, in

Chemical composition of tektites, obsidian, and Libyan Desert glass

Compound	Range in composition for tektites, %	Average for 73 tektites, %	Range in composition for obsidian, %	Average for 32 obsidians, %	Libyan Desert glass, %
SiO_2	68.00–89.81	74.64	66.70–78.15	70.40	98.20
Al_2O_3	6.13–15.42	12.12	11.54–18.30	15.26	0.70
Fe_2O_3	0.00–2.25	0.60	0.09–3.92	1.55	0.53
FeO	0.89–6.46	4.20	0.08–2.87	0.57	0.24
MgO	0.22–4.05	2.03	Trace–1.08	0.29	0.01
CaO	0.04–3.92	2.45	0.28–2.67	1.11	0.30
Na_2O	0.10–2.46	1.22	0.67–6.50	4.02	0.33
K_2O	0.82–3.60	2.23	1.45–5.38	4.24	0.02
TiO_2	0.18–1.40	0.82	*	*	0.33
H_2O	0.001–0.01	*	0.8–1.0	0.2	0.06

* Data not acceptable for averaging

the Philippine Islands, and in the lands adjacent

cumulating or were made within them by impacts of meteorites

Origin. The form and distribution of tektites suggest that small masses of viscous material entered our atmosphere, or formed therein before falling to the earth. How or where tektites originated is a mystery. Theories proposed are (1) tektites represent siliceous material melted from stony meteorites passing through the earth's atmosphere; (2) tektites are fused material thrown from the moon after either a meteorite or a comet collided with that body; (3) they were formed by a meteorite striking the earth; (4) they are fragments of an exploded planet; (5) tektites are the debris of a comet which passed close to the sun; (6) they were formed by the fusion of earth material when a comet collided with the earth's atmosphere.

If tektites are extraterrestrial in origin, several problems must be resolved: No falls have been witnessed; there are no intermediate varieties between tektites and meteorites, and tektites are apparently confined to a limited area, whereas meteorites fall in all regions of the earth. Since tektites contain about 100 times more uranium than stony

challenges the extraterrestrial origin of tektites. See METEORITE; ROCK (AGE DETERMINATION).

[E.P.H.]

Telecast

A television broadcast, involving the transmission of the picture and sound portions of the program by separate transmitters at assigned carrier frequencies within the 6-megacycle-wide channel assigned to a television station. A telecast is intended

modulation, while the sound portion (in the United States) uses frequency modulation. The channels assigned for telecasts by the Federal Communications Commission, each 6 megacycles (Mc) wide, cover frequencies as follows: 54–72 Mc (channels 2 through 4), 76–88 Mc (channels 5 and 6), 174–216 Mc (channels 7 through 13), and 470–890 Mc (channels 14 through 83). See TELEVISION.

[J.M.R.]

Telegraphy

A method of communication employing electrical signaling impulses produced and received manually or by machines. Telegraph signals are transmitted over open wire or cable land lines, submarine cables, or radio. Telegraphy as a communication technique uses essentially a narrow frequency band and a transmission rate adapted to machine operations. See COMMUNICATIONS, ELECTRICAL; DATA TRANSMISSION.

Early equipment devised by Samuel F. B. Morse consisted of a mechanical transmitter and receiver or register. Operators soon learned to handle messages faster using simple manual keys and audible sounders. Subsequently, telegraph transmission and reception again became mechanized. See TELETYPEWRITER EXCHANGE (TWX) SERVICE; TELEX. Telegraphy may also be used in other ways. See FACSIMILE; TELEPHOTOGRAPHY; TELETYPESETTER.

Telegraph facilities for use by the general public to transmit messages both domestically and internationally are provided by communication companies and government administrations. Special telegraph facilities include those for news services, distribution of market prices of securities and commodities, and private lines betweens such points as the factories and offices of a company for the exchange of messages, orders, payroll data, and warehouse inventories. Municipal and private fire and police alarms are a special form of telegraphy. The armed forces have extensive fixed and mobile telegraph systems.

Telegraph codes. For manual operation, the the code consists of short dot and long dash signals (Fig. 1). The original Morse code also used various length spaces, the Continental code avoids them. On submarine cables, the dots and dashes are of equal length for most efficient use of the transmission characteristics and are distinguished by being of opposite electrical polarity.

Most automatic printing telegraph circuits, including American cable operation, use a code of five equally spaced signals or units per letter or other symbol perforated into a paper tape (Fig. 2). The presence or absence of current or current reversals during these intervals constitutes the distinguishing feature. When a perforated tape is used, it usually is driven by a sprocket running in holes between the second and third code unit. Machines translate automatic teleprinter code to cable code, cable code to automatic code or, for special applications, make other translations. For stock quotation systems and teletypesetter opera-

Continental code		Morse code	
alphabet			
A	· –	A	· –
B	– · · ·	B	– · · ·
C	– · – ·	C	· · ·
D	– · ·	D	– · ·
E	·	E	·
F	· · – ·	F	· – ·
G	– – ·	G	– – ·
H	· · · ·	H	· · · ·
I	· ·	I	· ·
J	· – – –	J	– · – ·
K	– · –	K	– · –
L	· – · ·	L	⸺
M	– –	M	– –
N	– ·	N	– ·
O	– – –	O	· ·
P	· – – ·	P	· · · · ·
Q	– – · –	Q	· · – ·
R	· – ·	R	· · ·
S	· · ·	S	· · ·
T	–	T	–
U	· · –	U	· · –
V	· · · –	V	· · · –
W	· – –	W	· – –
X	– · · –	X	· – · ·
Y	– · – –	Y	· · · ·
Z	– – · ·	Z	· · · ·
numerals			
1	· – – – –	1	· – – ·
2	· · – – –	2	· · – · ·
3	· · · – –	3	· · · – ·
4	· · · · –	4	· · · · –
5	· · · · ·	5	– – –
6	– · · · ·	6	· · · · · ·
7	– – · · ·	7	– – · ·
8	– – – · ·	8	– · · · ·
9	– – – – ·	9	– · · –
0	– – – – –	0	⸺
punctuation			
(.)	· – · – · –	(.)	· · – – · ·
(,)	– – · · – –	(,)	· – · –
(?)	· · – – · ·	(?)	– · · – ·
(:)	– – – · · ·	(:)	– · – · ·
(;)	– · – · – ·	(;)	· · · · ·
(-)	– · · · · –	(-)	· · · · · – · ·
(!)	– – · · –	(!)	– – – ·
)	· – – – – ·	(')	· · – · · – · ·
)	– · · – ·	(/)	· · – –
)	– · – – · –	(	· · · · · – ·
		)	· · · · · · · · ·
)	· – · · – ·	(")	· · – · – ·
		(")	· · – · – · – ·

{there are no space letters in the Continental code

{C, O, R, Y, Z and & are composed of dots and spaces
T is a short dash
L is a longer dash
Zero (0) is usually abbreviated to T

1. Continental code is commonly used for tele-
ph communication Morse code continues in use
a few land lines in United States and Canada

tion, a six-unit code is used to provide control of machine action. For data transmission or machine control a seven- or eight-unit code may be used

Telegraph circuits and equipment. A single circuit provides transmission in only one direction at a time. For transmission in both directions si multaneously, a duplex circuit is used. Multiplex (time-division) apparatus provides two, three, or more channels operable in both directions simul taneously over a single circuit. Carrier current techniques enable several circuits, each comprising one or more communication channels, to operate through the same wide-band wire, cable, or radio facility. *See* TRANSMISSION THEORY AND METHODS.

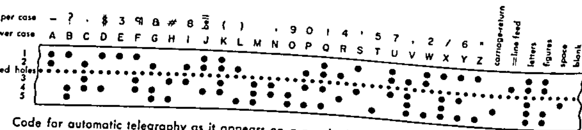
In automatic transmission, an operator at a manual keyboard, operated like a typewriter, perforates a tape. The tape is fed through an electro mechanical or photocell tape reader...

...communication channels carry these code pulses to the receiver. The received impulses may automatically actuate a reperforator to produce a duplicate punched tape for retransmission or later transcription; the impulses may actuate a teletypewriter (also called a teleprinter) to retype the original message; or they may actuate other terminal equipment, such as an accounting machine.

Alternatively, the equipment at both terminals may be teletypewriters, in which keyboard and printing mechanisms are combined in one machine. Such machines use a five-unit code but operate without perforated tape. Business firms that originate and receive numerous messages have such equipment installed at their offices for direct service. Also widely used are facsimile instruments that transmit or reproduce a typed or handwritten message as a picture on electrosensitive paper. [C H.]

Message service. Individual circuits are established as needed from a nation-wide network of central and branch telegraph offices and interconnecting wires, coaxial cables, and radio relay stations. The total traffic capacity of each of these interconnecting facilities may be divided by frequency-division multiplex into individual telegraph channels Such networks of national telegraph companies and government administrations are interconnected for world-wide service

Tie lines fan out from each central telegraph office to business, government, and other principal users of public telegraph services. Public telephones provide additional pick-up and delivery fa-



Code for automatic telegraphy as it appears on a punched tape (Western Union Telegraph Co.)

...lities for messages. Where traffic density is high, as in office buildings, messengers pick up and deliver telegrams. The trend for pick-up and delivery in the United States is toward facsimile because of its lower capital cost, less exacting requirements for manipulative skills, and error-free reception. In Europe the trend is toward TELEX, an automatically switched system with manually operated teletypewriter exchanges

Automatic switching. To achieve speed and accuracy in handling messages en route, direct circuits may be set up from the originating keyboard

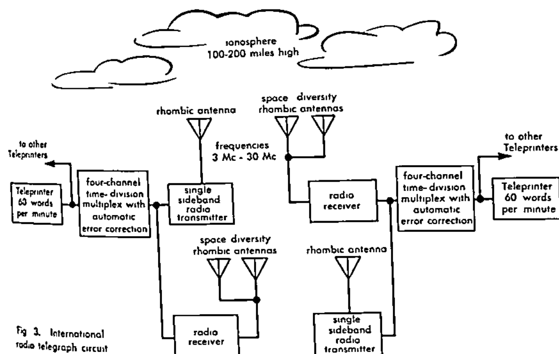


Fig. 3. International radio telegraph circuit

to the distant teleprinter. This is done in TWX and TELEX services Where direct channels are not immediately available, or where the volume of traffic calls for a storage interval, the message is transferred by means of perforated tape at switching centers. This method has the advantage of using the channel capacities of trunk lines more fully than with direct connections. It is used in the United States for public telegram service and on leased wire switched networks

As an example, follow a telegram from Providence, R I., to Walla Walla, Wash. In the Western Union telegraph system, the United States has 15 area switching centers, all directly and continuously interconnected. The message from Providence goes to the area switching center at Boston preceded by the director signal code PR perforated by the Providence operator. An automatic sensor at Boston receives this code (meaning Portland, Oreg.) actuates equipment to find an idle reperforator in the Portland trunk group and releases ...

signal its availability to a seeker switch, whose function is to connect reperforators with idle trunks. In this way the message will be transmitted on to Portland, where it will again be perforated and punched on tape.

The message will then appear before the Portland operator who, seeing the Walla Walla address, will route it by push button to its destination where it will be received on a teleprinter. [I.S.C.]

Overseas communication. In overseas communication, telegraph messages usually are referred to as cablegrams or radiograms, depending on the overseas transmission medium Private-line service is referred to as leased-channel service; customer-to-customer teletypewriter service, such as TWX in the United States, is termed TEX or TELEX or IMCO (International Metered Communication) in international service. The difficulty of providing frequent relays or repeaters in overseas services leads to extensive use of radio circuits operating at 3–30 Mc and of submarine cables.

Radio circuits Most radio circuits operate directly between countries (Fig. 3); only rarely is an intermediate relay used. Radio-telegraph trans-

[illegible]

accordance with the telegraph keying. In SSB transmission, the carrier is amplitude modulated by several subcarriers separated for frequency multiplexing, with one or two independent sidebands being transmitted with a reduced (suppressed) carrier; Power is usually 1–40 kw.

Radio receivers use either frequency- or space-diversity reception For frequency diversity, the

same telegraph keying is transmitted simultaneously on two different frequencies. For space diversity, two receiver antennas are spaced about 1000 ft apart. In either form of diversity reception, an automatic gate in the dual receiver selects the stronger signal. Various types of highly directive antennas are used, the rhombic being the most popular for both transmission and reception.

Radio waves at these frequencies are reflected between the ionosphere and the surface of the earth. Ionospheric characteristics vary from day to night, from season to season, by position on the earth's surface, and during the 11-year sunspot cycle (*see* IONOSPHERE). The transmission frequency is chosen in accordance with prevailing ionosphere conditions.

Most radio circuits operate with 60-word-per-minute (wpm) teletypewriters using the five-unit start-stop code; a few operate with terminal equipment using the Continental code. Multiple channels are obtained by using either time division or frequency division. Automatic error correcting equipment, called ARQ is used with two- or four-channel time-division multiplex to achieve high reliability. These radio circuits operate between all types of terminal equipment.

Cable circuit. The first cable between Europe and North America was used in 1858. Early submarine cables consisted of a center copper conductor, usually comprising 7 strands of No 18 BWG wire, wrapped with copper tape for continuity if the center strands failed. Guttapercha insulation, jute, armor wires, and an outside jute or braid covering completed the cable. Speeds up to 17 wpm were achieved. Submerged vacuum-tube repeater amplifiers, located about 200 miles offshore, and terminal equipment with signal-shaping amplifiers and time-division multiplex increased the traffic rate to six 50-wpm teletypewriter channels. The use of Permalloy for continuous cable loading further improved capacity to eight such channels Both the old and the new type cables are still in operation (Fig. 4)

Improved insulation and vacuum tubes with lives of 20 years or more have expanded the use of submarine cables. A new transoceanic telephone cable with submerged repeaters about 37 miles apart between North America and England has a capacity of 36 voice channels, each one of which can carry forty-eight 60-wpm telegraph channels by means of combined frequency- and time-division multiplex. These cables are for one-way nonreversib operation and are, therefore, laid in pairs. Eac consists of a solid central wire 0 13 in. in diamet surrounded by a spiralling copper tape, polyeth ene insulation, return copper tapes, jute, ar armor. [EDB

Private-line telegraphy. Organizations with se eral widely separated offices can communicate re ularly over lines leased from a common carri Various bandwidths for use with telegraph instr ments, teletypewriters, and various data handli terminal equipment are available. Companies th maintain their own rights of way, such as railroa and pipelines, may own their intercity wire lin

The larger private-line systems consist of one more relay centers each serving several tributa stations. This layout reduces the line mileage r quired to interconnect many tributary stations ha ing a known traffic volume. At a relay center me sages are received by a printer reperforator, whi prints the message and also perforates it in a tap At a small relay center an operator carries the ta to the proper outgoing transmitter for onward ha dling; at a large relay center the reperforated ta is automatically routed to the appropriate outgoi line in accordance with a 3–7 letter code address the beginning of each message.

A new application of private-line telegraphy lustrates the sort of communication that it pr vides. Hotel chains, and airline or railroad tick offices are interconnected by private telegraph lin that tie into a central file computer. The comput can be queried by local subsets about the size of desk adding machine The computer contains po ings of all available accommodations. If the co puter indicates that the desired accommodation available, operation of a key at the subset recor it as sold, the computer reduces its posting by on and the ticket agent completes the sale.

Accuracy of transmission is maintained by se eral methods. In a parity check for errors, a r dundant check pulse is sent with each character required to produce an odd number of mark pul Failure to receive an odd number of marks act ates an alarm signal Another technique is to a sign each of the pulses in a five-unit teleprint code a binary value: 1, 2, 4, 8, and 16 respectivel The sum of all binary values in an arbitrary bloc such as every 50 characters, is transmitted at th

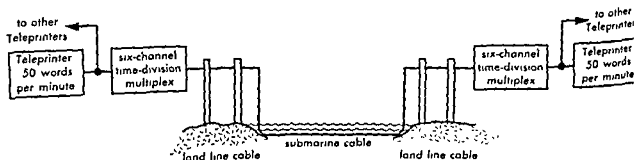


Fig. 4. Simplified transoceanic telegraph cable circuit.

of the block. If the transmitted sum checks the computed at the receiver, all is well; if not, receiver automatically requests a rerun of the orated tape. If the second transmission fails to k, an alarm is signaled [H M S.]

ock quotation ticker service. One of the older isive telegraph services is stock market quota- transmission This is a "broadcast" service t to a large number of receiving stations) orig- ing at a printer telegraph on the exchange . Abbreviated reports of stock sales on the are passed to operators who perforate tapes immediate transmission both directly to nearby kerage offices and through repeating relays at graph centrals to distant offices, the repeating ys and regenerative repeaters being spaced ely enough to operate on reliable signals not ously distorted by lengthy transmission cir- s.

ers a minute, receiving 8 pulses for each char- ter. The first pulse starts the ticker, the next six lses (intelligence pulses) position the wheel, the t or rest pulse stops the action (Fig 5) If the

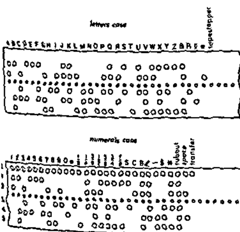


Fig. 5 Telegraph intelligence code for printing stock quotation systems.

rest pulse lasts longer than normal, the tape drive motor is also stopped. Letter characters are on the upper half of the wheel and numerals on the lower, printing in corresponding positions on the tape Where quotations are projected on a screen, transparent tape is used [P L.M]

Press systems Vast teletypewriter networks are used by the press associations and other news distributing organizations Most of these are modern carrier telegraph circuits operating over both wire and radio facilities.

Within the United States the press networks transmit to as many as 2000 receiving stations simultaneously Most of these networks operate at a speed of about 60 wpm, although considerably higher speeds (up to 1000 wpm) are now becoming available. On the average, about 1,000,000 words of news copy per day are processed through the New York offices of the major press associations. To handle this traffic as many as 40 circuits, having a total length of over 1,000,000 miles, and over 12,000 teletypewriters are in operation during peak operation. [V.N.V]

Telemetering

The branch of engineering concerned with the presentation of measured data at a location remote from the source of the data; also called telemetry. An example of a highly complex telemetering system is the equipment utilized to measure the temperature in a space vehicle in flight, radio this temperature to a ground observing station, and provide a meter or visual display to the observers. Telemetry encompasses all remote metering, whether the distances involved be many miles, as in space flights, or only a few feet, as in the measurement of reactivity in the core of a nuclear reactor See OPEN-LOOP CONTROL SYSTEM.

Telemetry involves three separate functions: (1) generation of a signal (electrical or otherwise), which measures the pertinent physical variable and is in suitable form for transmission; (2) transmission of the information to the remote location, and (3) conversion of the data into a form appropriate for display, recording, or application to further data-processing equipment.

Telemetry systems are conventionally classified according to the transmission medium Although the information can be transmitted in a variety of forms (such as the brightness of a modulated light beam), the great majority of systems fall into three categories: mechanical, electrical, and radio telemetry

Mechanical telemetry, involving mechanical coupling between the points of measurement and of data utilization, is generally restricted to short distances because of the high attenuation of mechanical media, the low velocity of propagation (of sound, for example), the difficulty of constructing efficient and simple mechanical amplifiers, and the requirement of a continuous mechanical coupling medium throughout the transmission path. Direct mechanical linkages through shafts and gear trains are used for telemetering over distances of a few feet, fluid coupling (hydraulic or pneumatic) is used in industrial process control for telemetering over distances up to several hundred feet. For a discussion of such mechanical telemetering systems, see SERVOMECHANISM

The term electrical telemetry is utilized in contrast to radio telemetry and refers to wired telemetry systems in which the information is transmitted by variations of a voltage or current in the electric circuit Modern electrical telemetry systems often involve complex electronic equipment, as when the information is transmitted (over wires) in the form of a television picture. The oldest and

most extensive use of electrical telemetry is in electric power distribution systems to collect, at one central location, data indicating the distribution of loads about the system so that the various generators (at different locations) may share the load in the most economical fashion. In the modern distribution system, data on loading is converted to suitable form, transmitted to a central location, and converted into a form suitable for processing by the digital computer which has primary control over the system.

Radio telemetry, originally used to telemeter information from weather balloons to ground observers, employs electromagnetic radiation as the transmission means. During the 1930s it was used in the testing of manned aircraft and drones. As a basic element of missile and space technology, radio telemetry is used to obtain, from unmanned space vehicles, data on environmental conditions (temperature, air density, radiation density, bombardment by micrometeorites) as well as on performance of the vehicle itself (strain, temperature, vibration).

Electrical telemetry. Electrical telemetry, first widely used at the end of the last century with the

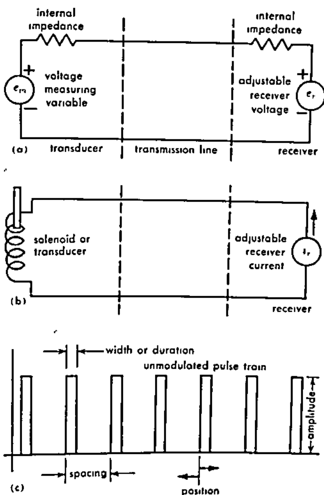


Fig 1. Electrical telemetry techniques. (a) Voltage-balance method; e_r adjusted to equal e_m. (b) Current or force balance; i_r adjusted to yield force that balances transducer force (c) Pulse telemetry, variables available for conveying information.

expansion of electric power systems and the railroads, today is also employed extensively in the supervision of such diverse systems as oil pipe lines and chemical process plants. Although a wide variety of types of systems are operative, the majority of systems fall into the following four categories.

Voltage-balance systems. At the measurement end of this type of system, a voltage is generated proportional to the variable to be telemetered The receiving equipment automatically generates a voltage to reduce the line current to zero (Fig. 1a)

Current-balance systems. The receiving equipment in these systems drives through the line a current just sufficient to yield an armature force counteracting the force corresponding to the vari

limited by loss, the presence of noise, and the slow response time of a long cable, and accuracy depends on these factors as well as the accuracy of balance.

Pulse systems. Pulse systems are used to minimize the effects of circuit parameters on accuracy. Information is transmitted by varying the number, amplitude, width, or spacing of a train of electrical pulses (Fig. 1c). In the simplest receiver, each pulse actuates a solenoid or motor, which moves the output one mechanical division.

Frequency systems In these systems the information is transmitted by modulating (varying according to the signal) the frequency or amplitude of the output of an electronic oscillator. In both these systems and the pulse systems, the accuracy of transmission is vastly better than in the balance systems Both systems can be used with multiplexing (sending several signals simultaneously over a single pair of wires), and the rate of response can be much greater The recent elaboration of the pulse and frequency systems is found in television telemetry in which the picture is transmitted by a sequence of pulses of varying amplitude and the sound by a frequency system. For a discussion of electrical telemetry methods, particularly for consideration of the techniques of information transmission with multiplexing, *see* TRANSMISSION THEORY AND METHODS

Radio telemetry systems. Radio telemetry, originated in Germany about 1930 for obtaining data from weather balloons is of primary interest today in missile engineering In this application, the typical telemetering system must transmit information on 30 or more variables simultaneously and with a low probability of error, a high equipment reliability, and with a transmitter equipment as simple, compact, light, and inexpensive as possible

The primary basis for classifying telemetry systems is the method employed to achieve the simultaneous transmission of the various signals, that is, the multiplexing. The primary radio signal is ordinarily a sinusoidal variation at a high frequency (for example, in the band from 216–220 or 2200-

00 Mc). The conversion of information from as many as 30 different channels to this frequency band is commonly accomplished as indicated in simplified form in Fig 2*a*.

The signals on channels 1–5 inclusive are assumed to be relatively slowly varying so that each can be adequately represented by periodic samples selected by a rotary, mechanical commutator. The signal on line A then consists of a sequence of pulses as shown in Fig 2*b*, with the first pulse representing the amplitude of e_1 at time t_1, the second pulse the amplitude of e_2 at time $t_1 + h$, etc. Thus, signal e_1 is replaced by a sequence of pulses, separated by time intervals of $5h$ seconds in this 5-channel example, e_2 is likewise represented by a train of pulses, each occurring h seconds after the corresponding e_1 pulse. The combination of a number of signals in this fashion on a time-sharing basis is termed time multiplexing. The representation of any one signal by a train of pulses, with each pulse amplitude proportional to the original signal amplitude at the time of sampling, is termed PAM—pulse amplitude modulation (in automatic control, the same signal is described as a sampled-data signal and the commutation is termed sampling). *See* SAMPLED-DATA CONTROL SYSTEM.

Thus, line A carries the 5 channels of information. If the samples are to represent each signal adequately, the sampling frequency, $1/(5h)$ cps in Fig. 2*b*, must be at least three times (and prefer-

...

ing, contact wear) to about 1000 pulses/sec, time multiplexing of N channels (with $1000/N$ pulses per second available for each channel) requires

...

variations (such as temperature in aircraft flight

...

... telemetered (as e_6 in Fig 2*a*) may contain high-frequency components. In both cases, it is common practice to utilize these signals to frequency modulate the sinusoidal, fixed frequency output of subcarrier oscillators (A and B in Fig 2). For example, if oscillator A in Fig 2 is stabilized at 10 kc, the output of modulator A is an FM (frequency modulation) signal, with instantaneous transmitted frequency varying about 10 kc according to the information signal. Oscillators A and B are operated at different frequencies, so that the output of the adder contains one frequency band carrying the information in channels 1–5, another band carrying the information of channel 6

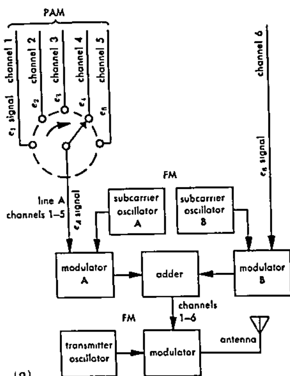


Fig 2 Typical transmitter modulation-multiplexing system (a) Simplified system block diagram. (b) Form of the PAM, time-multiplexed signal e_A

The total signal from the adder is used to modulate the primary transmitter. If frequency modulation is again used in this final modulator, the radiated output is an FM signal, centered about the carrier frequency. The instantaneous transmitted frequency carries the information of all input channels, 1–5 simultaneously with 6.

The system shown in Fig 2 involves two distinct multiplexing operations: the time multiplexing generated with the commutator, and the frequency multiplexing effected by the adder and modulator and accomplished as a consequence of the different subcarrier-oscillator frequencies. As a consequence of the double-multiplexing system, three distinct modulations are required, and the system is described as PAM-FM FM

The PAM-FM-FM system, although the commonest, is only one of a wide variety of possible schemes. Only two modulation stages may be required, particularly if all information channels are wide-band (FM-FM system used) or if all are narrow-band (PAM-FM employed). In addition, rather than FM in either of the last two stages, it is possible to employ AM (amplitude modulation) or PM (phase modulation). Similarly, the PAM stage might be replaced by PDM (pulse duration modulation), PPM (pulse position modulation), or PCM (pulse code modulation, in which the sample amplitude is converted to a binary number and transmitted as a train of on-or-off pulses), but usually the PAM system results in the cheapest, simplest, and most reliable modulation equipment (*see* PULSE MODULATION). In general, the multiplexing (in frequency, time, or both) of the information signals and the conversion to the frequency band available for radio transmission permits wide flexibility in the choice of the modulation scheme. For a discussion of the characteristics of different types of modulation, *see* MODULATION.

The choice of multiplexing and modulation schemes is a critical factor in the determination of the performance characteristics and accuracy of the telemetry system. Factors affecting the selection include the following:

1. The bandwidth required For the radio signal, the wider bandwidth of FM is utilized to reduce the effects of noise arising in the transmission and reception. Therefore lower transmitted power is required for signal detectability, and smaller, lighter, more economical transmitters are possible For the first stage, PAM (rather than PDM, PPM, or PCM) is almost always used because of the smaller bandwidth

2. The threshold, or the minimum required received-signal strength This factor is influenced by the noise in the communication system, by the difficulties of maintaining proper transmitter orientation (for example, satellite attitude) for maximum strength of the received signal, and by the variability of electromagnetic propagation conditions as the transmitter changes location. This factor is particularly important when the radio transmitter is to be designed for minimum weight, space, power, and cost.

3. The information efficiency, or the effectiveness with which the bandwidth is utilized to decrease the effects of the noise introduced in transmission and reception. *See* INFORMATION THEORY

4. Crosstalk, or interaction among the various information channels The modulator of Fig 2, for example, is a nonlinear device, yielding an output with components at the various sum and difference frequencies of the input. When many channels are being telemetered simultaneously, the many sum and difference frequencies of the various input components lie within other channels Thus, crosstalk limits the number of channels utilized in frequency multiplexing (with 24 a typical maximum number).

5. Practical difficulties associated with the construction of transmitters exhibiting linearity (to minimize not only crosstalk, but also nonlinear distortion). The simplicity of FM transmitters in this regard is a primary factor in the frequent selection of frequency modulation (or phase modulation) for the final stage.

6. The effects of shocks or vibrations on the performance of the transmitter (particularly troublesome in FM transmitters, where the microphonics result directly in variation of the transmitted frequency).

7. The complexity of the required equipment, particularly as it affects cost, weight, size, reliability, and ease of maintenance.

8. The flexibility of the system, especially as it influences design of the receiving equipment and preparation of the received data for further processing (as in a digital computer) or for recording (as on a magnetic tape) for later processing.

Telemetry equipment. Equipment used in telemetry is similar in most respects to electronic communication equipment. Certain features, however, are characteristic of telemetry applications.

Commutator. The receiver also must include a commutator (or time filter) to separate the time-multiplexed channels. Synchronization of the receiver commutator with that in the transmitter can be accomplished on the basis of pulse frequency or, more conveniently, by utilization of one channel for transmission of a synchronizing signal.

Subcarrier oscillators and modulators. In Fig 2, subcarrier oscillator A is frequency modulated by the electrical signal e_1, oscillator B, by the information signal e_6. The latter modulation can often be accomplished directly by the inclusion, in the oscillator or multivibrator, of a variable element or transducer responding to the variable to be telemetered For example, thermistor (or other temperature-sensitive) elements in *RC* oscillators permit variation of oscillator frequency directly with temperature; the use of variable inductances or saturable reactors in *LC* oscillators permits direct variation of frequency with mechanical position; strain-gage bridges can be used to adjust the feedback in *RC* oscillators to yield a frequency that directly measures strain (and hence force, stress, or torque) Regardless of the modulation method, a primary limitation on the FM-FM telemetry system is the requirement that the subcarrier frequencies be accurately stabilized—not only to avoid crosstalk, but more importantly to realize a high accuracy in the telemetering operation, since random frequency variations are added directly to the information frequency modulation

Transmitter and receiver. Conventional communication equipment is utilized (such as Armstrong or reactance-tube FM modulators) However, in air-to-ground telemetering, strong emphasis is placed on minimizing weight, size, cost, and power requirements.

Demodulators. The accuracy with which demodulators can be constructed is an important factor

selection of modulation and multiplexing es. The receiver corresponding to the typical of Fig. 2 includes filters to separate the in-tion in channel 6 from that in 1–5. Such fil-re widely used in communication systems (*see* ELECTRIC) and yield faithful reproductions original signals except for the following con-s:

Noise introduced in instrumentation, in com-cation, and in the low-signal-level stages of the er Noise is minimized by proper design, se-n of bandwidth, and choice of modulation acteristics.

Nonlinear distortion in the modulators and de ulators (particularly the final transmitter mod-r). Nonlinear distortion frequently can be at

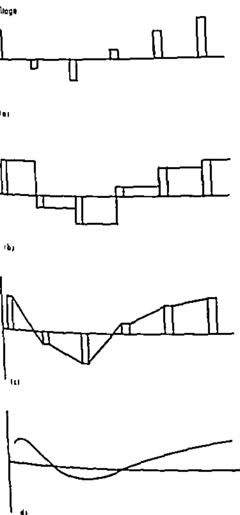


Fig. 3 Methods of PAM demodulation (a) PAM sig-nal for one channel (b) Hold circuit, value held until arrival of next pulse (c) Linear variation between samples (in actual system, output delayed by fixed amount 5h since linear curve cannot be generated until both end points are known) (d) Output when electrical network filter used to pass only low-fre-quency components of PAM signal

least partially compensated by nonlinear circuits operating on the demodulated signal

3. Amplitude and phase distortion introduced in transmission and amplification or in the filters. Such distortion can be removed (with a constant time de-lay resulting) by appropriate compensation net-works.

Demodulation of the PAM signal (after the re-ceiver commutator separates channels 1–5) is ac-complished with interpolation or holding circuits or with low-pass filters. Figure 3 illustrates the common schemes for recovery of a smooth signal from a sequence of samples. In the hold circuit, the value of one pulse is held until the arrival of the next pulse. To obtain linear variation between sam-ples as shown by Fig 3*c*, the output must be de-layed by a fixed amount equal to the time between pulses, since a linear curve cannot be generated until both end points are known.

Thus, the science of radio telemetry, of funda-mental importance in both industrial control and astronautics, involves electronic and electromag-netic communication systems incorporating com-plex modulation and multiplexing as well as the in-strumentation associated with both the measure-ment of physical quantities in electrical terms and the processing and storing of the data received.

[J.G.TR.]

Bibliography: M. H Nichols and L. L. Rauch, *Radio Telemetry*, 2d ed , 1956; *Proc. IRE* and *IRE Trans. on Space Electronics and Telemetry.*

Telephone

The term telephone was formerly applied to the telephone receiver, the instrument originally in-vented by Alexander Graham Bell. The term is now commonly applied to the telephone set, which in-cludes a transmitter and an electric network in addition to the receiver.

The transmitter and receiver are housed together in a handset A cord connects the electrical com-ponents of the handset to the network in the tele-phone set.

Transmitter. The transmitter is a transducer which converts acoustical energy into electric en-ergy In most transmitters an electric current is modulated by the variations in contact resistances of carbon granules Sound waves impinge on the diaphragm of the transmitter, causing the carbon granules to move closer together, making more contacts and decreasing resistance, or to move further apart, making fewer contacts and higher resistance.

The transmitter has a frequency response range from 250 to 5000 cycles per second (cps). The frequency response rises uniformly to a broad maximum in the region of 2500 cps Because of these characteristics and those of the telephone receiver, the speech heard by the listener resembles closely that of direct mouth-to-ear speech as heard by a listener a few feet from the person speaking.

The transmitter is of the direct-action granular-carbon type shown in Fig. 1. It consists essentially

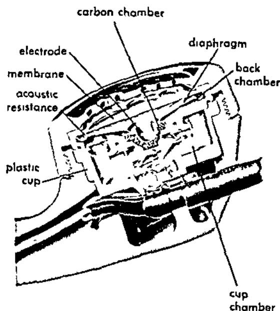


Fig 1. Direct action granular carbon telephone transmitter.

of a diaphragm, back cavity, and carbon chamber. The diaphragm is rigidly clamped at its periphery to obtain a high output in the upper frequency range This achieves a quality of transmission that approximates the orthotelephonic. mouth-to-ear transmission through the air objective. The sound pressure on the diaphragm varies the pressure of the dome-shaped electrode on the carbon. Changes in pressure on the carbon granules vary the resistance of the granules, causing the current to change in magnitude in proportion to the sound. The carbon granules are specially treated to reduce the effect of aging on their resistance The carbon chamber is designed to keep the mechanical impedance of the carbon to a minimum, regardless of the position of the transmitter unit, to gain a high modulating efficiency.

The desired frequency response is obtained by coupling the diaphragm to a doubly resonant system, composed of a cavity within the unit behind the diaphragm and a chamber between the unit and a plastic cup. The two cavities are connected by holes covered with a woven fabric The size of the hole and resistance of the material to the flow of air is carefully controlled to balance the acoustical impedance and therefore prevent large irregularities in the transmitter response.

The transmitter assembly is located in a specially designed handset also housing the receiver.

Receiver. The receiver transducer operates on the relatively low power, used in the telephone circuit, to convert electric energy into acoustical energy. Unlike a loud-speaker, the telephone receiver is designed for close coupling to the ear The telephone receiver has an approximate impedance of 150 ohms at a frequency of 1000 cps. The relationship of the acoustic and electrical elements produces a desired response-frequency characteristic.

There are two types of receiver units, the bipolar receiver and the ring armature receiver shown in Fig. 2. The bipolar receiver is used with operator's head telephone sets. The ring armature receiver is used with the telephone set where a high efficiency is particularly important.

The advantages of the ring armature receiver are its low acoustic impedance and high available power response over a wide frequency range (350-3,500 cps). These advantages are achieved by the piston action of a thin, nonmagnetic, light weight, dome-shaped diaphragm, which is attached to a ring-shaped armature of magnetic material driven at its periphery by a ring-shaped magnetic coil associated with a ring-shaped permanent magnet The diaphragm contains a small hole that introduces a low-frequency cut-off. This is desirable to reduce interference picked up from electric power circuits.

The diaphragm, magnets and coil are encased in a ferrule grid attached to a molded terminal plate A membrane between the ferrule grid and diaphragm protects the diaphragm from dirt and mechanical damage Because of its mechanical impedance, the membrane acts as one of the controls over the diaphragm to help achieve the desired frequency response

The acoustic chamber between the membrane and diaphragm is connected to a chamber molded in the terminal plate behind the diaphragm, called the back chamber. These chambers are connected by passageways having acoustic mass and resistance The back chamber exhausts into the handset through acoustic fabric. All these controls are designed to extend the frequency range of the receiver and reduce undesirable diaphragm resonance. A click-reducing varistor is mounted on the back of the receiver

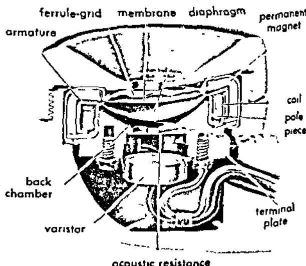


Fig 2 Ring armature telephone receiver.

Electric network. The electric network serves three basic functions: (1) to couple the receiver to the transmitter circuit, (2) to balance the impedances within the telephone set to those of the circuit and reduce sidetone and (3) to provide for transmission equalization, that is, the same general loudness and quality of speech over any circuit between the telephone set and the central office.

detone balance Sidetone is the sound of the

his voice and is therefore objectionable To p sidetone at a minimum, any voltage developed the local transmitter is divided in the windings and B (see Fig. 3) so that the voltages induced winding C are opposing. Also, the voltage across network resistance arising from the current wing in winding B opposes the resultant voltages luced in C. The over all effect is that the current the receiver, as a result of voltages *developed in* transmitter, is small. However, the key to good letone balance is to balance effectively the impedances Z_1 and Z_2 both in magnitude and phase *will vary because it is influenced by the tele-* phone circuit to the central office The essential ements in the impedance matching are the two licon carbide varistors V_1 and V_2 The dc and ac sistances of these varistors vary with the voltages pplied to them, which are in turn dependent on the irect current in the loop and in the telephone set his current is a function of the length and impedance of the circuit

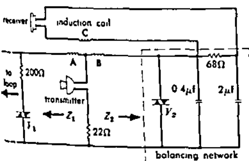


Fig. 3 *Transmission circuit* of a telephone set illustrating sidetone balance.

Transmission equalization. The transmission equalization provided by *the circuit assures that* the over all transmission performance of the set is kept within reasonable limits as the loop length varies. Equalization is provided by reducing the transmitting and receiving gain on short loops This is accomplished by a varistor V_1 across the

the shunting effect of the varistor decreases Thus the transmitting and receiving levels are reasonably constant for all length loops.

Accessory devices. In addition to the basic components, the telephone set has a switchhook to connect it to the line and operate other associated features, a dial which produces dc pulses to actuate the switching gear in the central office, and a ringer with an adjustable volume control, used to signal that a call is incoming. These are all enclosed in a common housing

A variety of optional features are provided for telephone sets. Transistor amplifiers are available to amplify speech at the transmitter to aid weak voices, to amplify received speech, to overcome ambient room noise or *hearing difficulties*, and to provide adequate transmission at reduced currents on extremely long loops. Telephone sets also are

on the customers' premises *See* INTERCOMMUNICATING SYSTEM; TELEPHONE SERVICE; TELEPHONE SIGNALING, SPECIAL SUBSCRIBER. [R F D]

Telephone civil defense system

A telephone system especially arranged or designed to assist in protection of civil population, limitation of casualties and property damage, and preservation of maximum civilian support of military effort. Civil defense requirements for communication fall into two broad categories, attack-warning communication, and administrative and control communications to alleviate disaster after an attack Both of these requirements are usually best met by private-line telephone systems furnished by the telephone companies, supplemented *by local* and interurban telephone service arranged to assure continuity of service *See* TELEPHONE SERVICE.

Attack warning. In the United States the basic *facility for attack warning is a private line tele-* phone system known as the National Warning System (NAWAS) This system connects nearly 300 warning points within the United States with four warning centers. Warning centers are located at important air defense installations so that full advantage can be taken of the intelligence information available at these installations. These centers are located on the East Coast, in the Central States, on the West Coast, and in the Rocky Mountain area (National Warning Center) At each of these four centers, a representative of the Office of Civil and Defense Mobilization can disseminate informa*tion to all warning points in the* center's normal area, advising that attack is on the way and when it can be expected. Should any warning center be rendered inoperative, adjacent centers can cover the warning points that depend on the *inoperative* center

The NAWAS network operates 24 hours a day and is designed to by-pass, insofar as practicable, major cities which are likely to be enemy targets Wire and microwave radio facilities are utilized to accomplish this objective.

From the 300 primary warning points on the NAWAS network, warning information is relayed *through state and local* systems to more than 3500 secondary warning points. In many states private-line systems, both telephone and telegraph, extend from primary warning points to the secondary warning points The telephone companies *offer* a simple, reliable signaling system, known as bell and lights, to transmit warning With the bell-and-lights system, a master primary station, by dialing over a

private-line network, can signal any one of four alert conditions to a large number of bell-and-lights secondary receiving stations simultaneously. At these secondary warning points, local authorities are responsible for alerting the general population. Sirens, horns, whistles, and similar devices are used for this purpose.

Administrative and control communications. Communications to alleviate disaster are basically provided through National Communications System (NACOM I). This is a network of approximately 20,000 miles of wire and radio facilities arranged for optional use for either telephone or teletypewriter communications. Telephone and teletypewriter terminal equipment for this network is in place and ready to operate. The network itself is provided by the telephone companies and is set up as occasion arises on call of the Office of Civil and Defense Mobilization (OCDM).

The NACOM I network radiates from OCDM headquarters at Battle Creek, Mich., and from an alternate headquarters, to each of OCDM's eight regional offices. From these regional offices the system extends to the state headquarters of civil defense in each state of the region. In an attack situation, the NACOM I network will be urgently needed for collection of damage and radiation-fall-out information, and for directing the flow of ma power and material to effect conservation and rap restoration of the nation's resources.

The telephone industry attempts to provide t utmost in service reliability for OCDM. Probal target areas are by-passed, where possible, and t or more access routes to the nationwide telepho network are often provided. Because of its int laced pattern, the network permits a great vari of patching arrangements to by-pass damaged p tions of the system. Since most telephone cables a underground and radio relay systems are subject damage only at terminal and relay stations, so resistance to attack is inherent. [J.A

Telephone private branch exchange (PBX)

A switchboard and associated equipment, usua located on customers' premises, to provide switching calls between any two extensions serv by the PBX or between any extension and the tionwide telephone system via a trunk to a cent office. A PBX also frequently includes tie trun for direct communication with another PBX servi the same customer at a distant location.

Manual PBX. All switching functions in a man PBX are performed by an attendant.

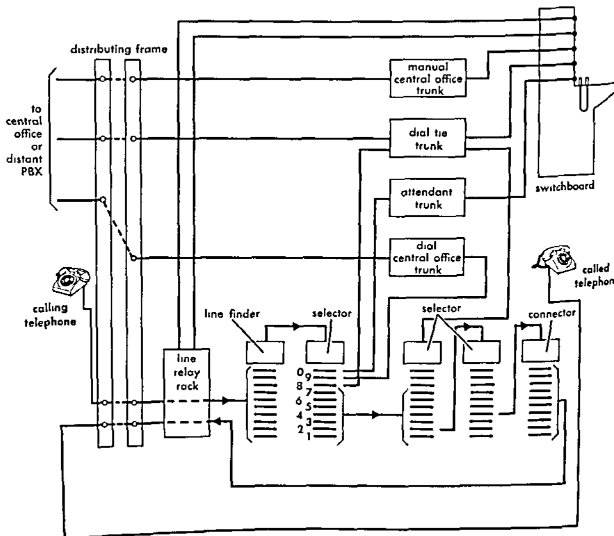


Diagram of step-by-step PBX

With a cordless manual PBX the operator uses keys to operate relays, which perform the actual switching functions. The attendant may use either an ordinary telephone or an operator-type headset.

The cord type manual PBX has jacks, each with an associated lamp and designation, for each extension, central office trunk and tie trunk. Cord circuits terminated in pairs of flexible cords tipped

...

priate keys.

The multiple manual PBX differs from the nonmultiple in that each extension or trunk is connected to more than one jack, so that it can be reached by any one of several attendants This facilitates prompt answers, particularly during periods of heavy traffic. The larger PBXs are generally of the multiple type.

Dial PBX. All extensions of a dial PBX are equipped with dials Electromechanical or other remote switching equipment is provided to make the desired connections Most dial PBXs also include an attendant's switchboard

The switching equipment may be of any suitable type. The illustration shows a schematic diagram of a typical dial PBX using step-by-step equipment Extensions, central office trunks, and tie trunks are connected to the PBX equipment via a distributing frame, for flexibility of assignment Extensions are each cross-connected to a line circuit, usually containing a line relay and a cut-off relay, and thence to the banks of a group of line finders. Each line finder is permanently connected to a particular first selector. When a call is originated at an extension by lifting the handset the line circuit calls for an idle line finder, which hunts over its associated bank until it finds the calling line The corresponding first selector then returns a dial tone, to indicate that it is ready to accept dial pulses. When the extension user dials the first digit, the wipers of the first selector step upward to the level corresponding to the digit dialed, and auto-

...

Ninth level trunks usually provide access to the nearest central office The eighth level is usually assigned to tie trunks. Levels two to seven inclusive, generally provide access to other extensions, via additional stages of selectors and, finally connectors. See SWITCHING CIRCUIT, SWITCHING SYSTEMS (COMMUNICATIONS), TELEPHONE SERVICE

Power plants are usually installed on the customer's premises to provide the PBX with the needed types of current. These usually include a storage battery, with a rectifier or motor generator to charge it, for supplying direct current to relays, dial switching equipment, and switchboard lamps. Small PBXs may obtain their direct-current supply from the central office battery via feeder wires. Also, many small PBXs use batteryless rectifiers, except those serving such customers as hospitals and fire and police departments where a battery is usually provided to insure continuity of service in case of power failure. Ringing current (generally 20-cycle alt... In the larger ir... generator set. ... by feeder wires from the central office ringing generator. [R F.D]

Telephone service

The technology of supplying the many types of services offered to the telephone subscriber. These services include local, long-distance, and overseas connections, information services, answering services and the leasing of private lines. For other discussion of the equipment and means of providing these services, see SWITCHING SYSTEMS (COMMUNICATIONS); TELEPHONE; TELEPHONY.

LOCAL TELEPHONE SERVICE

A local telephone service operates within an exchange, or within a connected system of exchanges in the same exchange area, and furnishes to subscribers an intercommunicating service. An exchange is a unit established by a telephone company for the administration of telephone service in a specified area, usually a town, a city, or a village and its environs The exchange consists of one or more central offices together with the associated plant used in furnishing telephone service in that area. An exchange area is the territory served by an exchange.

To provide this intercommunicating service, centrally located switching equipment is used to establish a communication connection between any two telephones. This equipment also provides for connections to distant exchanges

Local exchange transmission circuits are usually two-wire cable pairs operated at voice frequency Trunk cable pairs between exchanges may employ repeaters or amplifiers In the longer transmission circuits, carrier systems are employed See TRANSMISSION THEORY AND METHODS

Registration of messages. The proper charging of telephone calls to the subscriber requires the recording of data on each call. Message registers and automatic message accounting are systems for recording these data A message register is a counting device located in the central office and associated with each customer's telephone The message units totaled on the message register are used as a basis for bulk-billing the customer for local telephone service. Automatic message accounting uses a punched-tape or magnetic-tape recording of the billing information pertaining to the telephone message. The recording equipment may be located in the local central office or at a tandem or long-distance office.

Design of system. The design of a telephone system involves a consideration of telephone sets, subscriber loops, telephone central offices, interoffice trunks, and numbering plans.

Telephone set. Telephone sets may be of the noncoin- or coin-operated types. They may be manually or dial operated. A main telephone set, together with extension telephones, may have exclusive use of a telephone line. This is individual-line service. Sometimes the use of a single telephone line is shared by two, four, or more customers. This is party-line service. Many telephone lines terminate on private branch exchanges to service a large number of extension telephones. *See* TELEPHONE PRIVATE BRANCH EXCHANGE (PBX).

Subscriber loops. These are the relatively short circuits, or lines, connecting the customers' telephone sets to a central office. These lines usually employ 26-, 24-, and 22-gage copper wire. The efficient modern telephone set and central-office equipment permit the use of a high proportion of fine-gage wire. Resistance is usually the limiting factor in the selection of wire gage for subscriber loops.

Telephone central office. The central office houses the switching equipment that serves the telephones of its immediate community. The telephone building, housing one or more local central offices, is the "wire center" at which the customers' loops of the telephones in the vicinity converge and are terminated. The primary function of the central office is to initiate and control the switching action which, on demand, will permit the connection of any local telephone to any other telephone served by the same central office, in the same exchange, in nearby exchanges, and in distant cities For this purpose, each central office has direct interoffice trunks radiating from it to many of the other local central offices in the same area, to tandem offices and to long-distance offices for intermediate connection to circuits to more distant places.

A tandem office is an intermediate switching location whose primary function is to interconnect trunk circuits to and from the local central offices in the area In this manner trunking economies are achieved, particularly in those instances where traffic volumes between two local central offices are not sufficient to justify a considerable number of direct interoffice trunks between the two local offices.

Interoffice trunks These are the speech-carrying telephone circuits between the switching systems in the exchange and its environs. Carrier systems usually provide the transmission medium for the longer trunk circuits. For shorter trunk circuits, under about 8–10 miles, the two-wire circuit operating at voice frequencies continues to be the basic transmission medium. These two-wire circuits are 19-, 22-, or 24-gage, with some 26-gage cable pairs Transmission considerations usually indicate a requirement for inductive loading, and 88-millihenry coils spaced at 6000-ft intervals are commonly used A large percentage of these trunks are equipped with negative impedance-type repeaters either at the terminals or intermediate points Transmission requirements and resistance limitations control the selection of wire gage, loading, and repeaters. Signaling on virtually all voice-frequency trunks is by direct current. Trunk resistance limitations vary between 600 and 3000 ohms, depending on the types of central office involved.

An assemblage of interoffice trunks, used to carry a parcel of telephone traffic between two switching systems, is called an interoffice trunk group. The number of trunks in the group is dependent on (1) the probable amount of telephone traffic constituting the offered parcel during the busiest hour and (2) the quality of service to be maintained. In final routes, the number of trunks in the group is such that the offered traffic finds little or no delay in completion during the busy hour. In high-usage routes, the number of trunks is deliberately fixed at less than the number required for final service. Therefore, some of the offered traffic finds no idle trunks during the busy hour and is rerouted automatically, where the switching system permits, to another route via a tandem office This principle is termed alternate routing The rerouting involves no noticeable delay, and since the rerouted traffic is combined with other rerouted traffic items, a more efficient interoffice trunk network is obtained. *See* COMMUNICATIONS SYSTEMS (TRAFFIC) DESIGN.

An exchange cable is a grouping of copper conductors insulated by either paper pulp or polyethylene, covered by an outer sheath of protective material, and used for subscriber loops and interoffice trunks. Wire sizes commonly employed are 26-, 24-, 22-, and 19-gage. Cable sizes vary from 6 pairs to 2100 pairs. In the modern polyethylene-insulated cable (PIC), the cable is made up in 25-pair groups of conductors, each pair of which is identified by a distinctive color code. Each 25 pair group also has a distinctively colored binder to aid further in locating a particular pair. The protective outer sheath is usually made of plastic.

Numbering plans The numbering plan must uniquely identify each telephone station so that calls may be directed to it For a complete discussion, *see* SWITCHING SYSTEMS, COMMUNICATIONS

INTERURBAN TELEPHONE SERVICE

Interurban service provides and operates the equipment for switching and transmitting telephone communications between exchange areas. The term long-distance service is often used to denote interurban telephone service

Long-distance office. A centralized location in an exchange area is provided to handle long distance service for that area and often for other nearby exchange areas Large urban areas may have more than one subsidiary long distance office when it becomes necessary because of physical limitations at a single office or to protect service by decentralization. Figure 1 shows how long-distance offices handle interurban telephone calls originating and terminating in the exchange areas served.

Calls between exchange areas served by two different long-distance offices are often too few in

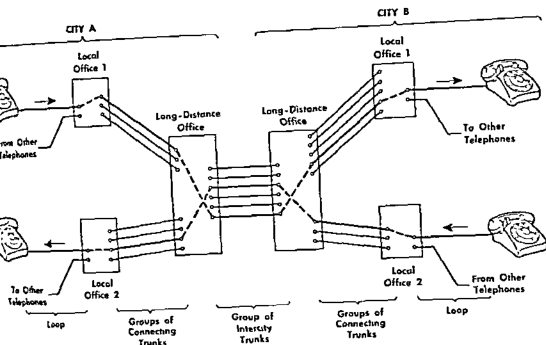


Fig. 1 Interconnection for long-distance telephone calls. Arrows indicate direction in which connection is established. (*American Telephone and Telegraph Co*)

number to justify economically a direct connection between the two long distance offices. Such calls are routed through a third long-distance office, which can connect directly to the first two offices. At the third long-distance office, these calls are termed through-switched traffic. Connections between offices are provided by groups of trunks. The number of trunks engineered for a group depends on the number of calls it is expected to carry.

The equipment at a long-distance office may be manual or mechanized. Switchboard positions at which long-distance operators establish connections by verbal request and write tickets to record billing information are used at a manual office. Most long-distance offices have some switchboard positions for operators who render assistance on interurban calls which require special handling or which cannot be completed by dialing procedures. Interurban telephone service is highly mechanized. More than 90% of the intercity trunks in the United States are equipped for the dialing of calls to completion by either operators or customers. Mechanized offices are equipped with dial-switching systems designed for interurban service. The type of switching system used depends on the num-

equipment needed for trunks and signaling.

Intercity trunks. These are the telephone channels between exchange areas. The distortion and attenuation limits are kept within a narrow range, since these effects are cumulative for the trunks that comprise a connection. At the minimum, there is a connecting circuit between the long-distance office and a local office at each end of an intercity trunk. If an interurban call is between two small towns 3000 miles apart, several intercity trunks in tandem may be used for the connection.

Many types of transmission systems are used for intercity trunks. For a particular long distance route, the most economical type is used for the distance to be covered and the total number of trunks to be provided over a reasonable length of time, that is, future growth must be considered. Transmission systems used include voice-frequency circuits on wire cable or open-wire lines, and carrier systems for wire cable, open-wire lines, coaxial cable and radio relay. Intercity trunks may be one-way, over which a telephone call can be initiated from one end only. Of course, conversation may be in either direction once the connection is made. Over two-way intercity trunks, a call can be initiated from either end, although not simultaneously.

For each intercity trunk, signaling equipment is provided to pass supervisory signals needed to establish a connection, to disconnect, to indicate a busy telephone line, and to indicate when all circuits are busy.

Recording of billing information. On most interurban telephone service, information is recorded in order to bill the customer correctly. There are three categories of billing information for interurban calls: (1) calls which are included in the charge for local service, (2) calls for which the total charge is a multiple of a unit charge; the total number of units used in a billing period are lumped together as an aggregate or bulk item; multiunit calls are generally recorded automatically, (3) calls which are billed individually with

Telephone set. Telephone sets may be of the noncoin- or coin-operated types. They may be manually or dial operated. A main telephone set, together with extension telephones, may have exclusive use of a telephone line. This is individual-line service. Sometimes the use of a single telephone line is shared by two, four, or more customers. This is party-line service. Many telephone lines terminate on private branch exchanges to service a large number of extension telephones. *See* TELEPHONE PRIVATE BRANCH EXCHANGE (PBX).

Subscriber loops. These are the relatively short circuits, or lines, connecting the customers' telephone sets to a central office. These lines usually employ 26-, 24-, and 22-gage copper wire The efficient modern telephone set and central-office equipment permit the use of a high proportion of fine-gage wire. Resistance is usually the limiting factor in the selection of wire gage for subscriber loops.

Telephone central office. The central office houses the switching equipment that serves the telephones of its immediate community. The telephone building, housing one or more local central offices, is the "wire center" at which the customers' loops of the telephones in the vicinity converge and are terminated. The primary function of the central office is to initiate and control the switching action which, on demand, will permit the connection of any local telephone to any other telephone served by the same central office, in the same exchange, in nearby exchanges, and in distant cities For this purpose, each central office has direct interoffice trunks radiating from it to many of the other local central offices in the same area, to tandem offices and to long-distance offices for intermediate connection to circuits to more distant places.

A tandem office is an intermediate switching location whose primary function is to interconnect trunk circuits to and from the local central offices in the area In this manner trunking economies are achieved, particularly in those instances where traffic volumes between two local central offices are not sufficient to justify a considerable number of direct interoffice trunks between the two local offices.

Interoffice trunks These are the speech-carrying telephone circuits between the switching systems in the exchange and its environs. Carrier systems usually provide the transmission medium for the longer trunk circuits For shorter trunk circuits, under about 8–10 miles, the two-wire circuit operating at voice frequencies continues to be the basic transmission medium. These two-wire circuits are 19-, 22-, or 24-gage, with some 26-gage cable pairs. Transmission considerations usually indicate a requirement for inductive loading, and 88-millihenry coils spaced at 6000-ft intervals are commonly used. A large percentage of these trunks are equipped with negative-impedance-type repeaters either at the terminals or intermediate points. Transmission requirements and resistance limitations control the selection of wire gage, loading, and repeaters. Signaling on virtually all voice-frequency trunks is by direct current. Trunk resistance limitations vary between 600 and 3000 ohms, depending on the types of central office involved.

An assemblage of interoffice trunks, used to carry a parcel of telephone traffic between two switching systems, is called an interoffice trunk group. The number of trunks in the group is dependent on (1) the probable amount of telephone traffic constituting the offered parcel during the busiest hour and (2) the quality of service to be maintained. In final routes, the number of trunks in the group is such that the offered traffic finds little or no delay in completion during the busy hour. In high-usage routes, the number of trunks is deliberately fixed at less than the number required for final service. Therefore, some of the offered traffic finds no idle trunks during the busy hour and is rerouted automatically, where the switching system permits, to another route via a tandem office This principle is termed alternate routing The rerouting involves no noticeable delay, and since the rerouted traffic is combined with other rerouted traffic items, a more efficient interoffice trunk network is obtained *See* COMMUNICATIONS SYSTEMS (TRAFFIC) DESIGN.

An exchange cable is a grouping of copper conductors insulated by either paper pulp or polyethylene, covered by an outer sheath of protective material, and used for subscriber loops and interoffice trunks Wire sizes commonly employed are 26-, 24-, 22-, and 19-gage. Cable sizes vary from 6 pairs to 2100 pairs In the modern polyethylene-insulated cable (PIC), the cable is made up in 25-pair groups of conductors, each pair of which is identified by a distinctive color code. Each 25-pair group also has a distinctively colored binder to aid further in locating a particular pair. The protective outer sheath is usually made of plastic.

Numbering plans. The numbering plan must uniquely identify each telephone station so that calls may be directed to it. For a complete discussion, *see* SWITCHING SYSTEMS, COMMUNICATIONS

INTERURBAN TELEPHONE SERVICE

Interurban service provides and operates the equipment for switching and transmitting telephone communications between exchange areas. The term long distance service is often used to denote interurban telephone service.

Long-distance office. A centralized location in an exchange area is provided to handle long distance service for that area and often for other nearby exchange areas. Large urban areas may have more than one subsidiary long distance office when it becomes necessary because of physical limitations at a single office or to protect service by decentralization. Figure 1 shows how long-distance offices handle interurban telephone calls originating and terminating in the exchange areas served.

Calls between exchange areas served by two different long-distance offices are often too few in

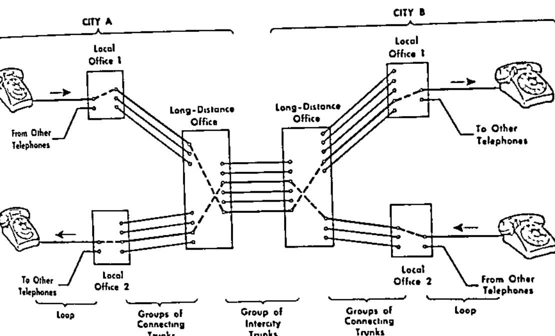


Fig. 1 Interconnection for long-distance telephone calls. Arrows indicate direction in which connection is established (*American Telephone and Telegraph Co*)

number to justify economically a direct connection between the two long-distance offices. Such calls are routed through a third long-distance office, which can connect directly to the first two offices. At the third long distance office, these calls are termed through-switched traffic. Connections between offices are provided by groups of trunks. The number of trunks engineered for a group depends on the number of calls it is expected to carry.

The equipment at a long-distance office may be manual or mechanized. Switchboard positions at which long distance operators establish connections by verbal request and write tickets to record billing information are used at a manual office. Most long-distance offices have some switchboard positions for operators who render assistance on interurban calls which require special handling or which cannot be completed by dialing procedures. Interurban telephone service is highly mechanized. More than 90% of the intercity trunks in the United States are equipped for the dialing of calls to completion by either operators or customers. Mechanized offices are equipped with dial-switching systems designed for interurban service. The type of switching system used depends on the number of calls to be handled and the distribution of traffic among the three categories of calls discussed above. Long-distance offices are also provided with the equipment needed for trunks and signaling.

Intercity trunks. These are the telephone channels between exchange areas. The distortion and attenuation limits are kept within a narrow range, since these effects are cumulative for the trunks that comprise a connection. At the minimum there is a connecting circuit between the long distance office and a local office at each end of an intercity trunk. If an interurban call is between two small towns 3000 miles apart, several intercity trunks in tandem may be used for the connection.

Many types of transmission systems are used for intercity trunks. For a particular long distance route, the most economical type is used for the distance to be covered and the total number of trunks to be provided over a reasonable length of time; that is, future growth must be considered. Transmission systems used include voice-frequency circuits on wire cable or open-wire lines, and carrier systems for wire cable, open-wire lines, coaxial cable and radio relay. Intercity trunks may be one-way, over which a telephone call can be initiated from one end only. Of course, conversation may be in either direction once the connection is made. Over two-way intercity trunks, a call can be initi-

tablish a connection, to disconnect, to indicate a busy telephone line, and to indicate when all circuits are busy.

Recording of billing information. On most interurban telephone service, information is recorded in order to bill the customer correctly. There are three categories of billing information for interurban calls. (1) calls which are included in the charge for local service, (2) calls for which the total charge is a multiple of a unit charge; the total number of units used in a billing period are lumped together as an aggregate or bulk item; multiunit calls are generally recorded automatically, (3) calls which are billed individually with

a listing of the locality or exchange called and the amount of the charge for each. These calls may be manually recorded by operators or recorded automatically.

When information for billing is recorded automatically, it is by registers assigned to individual lines for multiunit calls or by automatically printed tickets, punched paper tape, or magnetic tape for multiunit calls and detailed billed calls. In the United States, the Bell System uses registers, printed tickets, and punched paper tape for automatic recording. The latter equipment is known as automatic message accounting (AMA) equipment.

[illegible]	[illegible]	[illegible]	[illegible]
	regional center	RC	1
	sectional center	SC	2
	primary center	PC	3
	toll center	TC	4
	end office	EO	5

Fig 2 Standard routing pattern. *(American Telephone and Telegraph Co)*

Direct distance dialing (DDD). Inaugurated in the United States in 1951 and in Canada in 1958, this service permits a customer to dial his own interurban calls. DDD service requires dial-switching equipment for the local office at the originating end and all long-distance offices through which a DDD call may be routed. Automatic equipment is required to record the calling and called telephone numbers and the duration of the conversation. A comprehensive numbering plan is required to give each telephone a unique number to indicate the destination of a call to the dial switching systems. Finally, a standard routing pattern and a general switching plan are required. The standard routing pattern provides a basis for the selection of important switching centers, called control switching points (CSPs). The general switching plan provides an orderly method of interconnection over the most direct and economical route available

Standard routing pattern. The basic pattern is shown in Fig. 2. Each local office is called an end office Each long-distance office is classified as a toll center (TC), primary center (PC), sectional center (SC), or regional center (RC), in order of increasing importance or rank PCs, SCs, and RCs are all CSPs. Dial-switching equipment with the

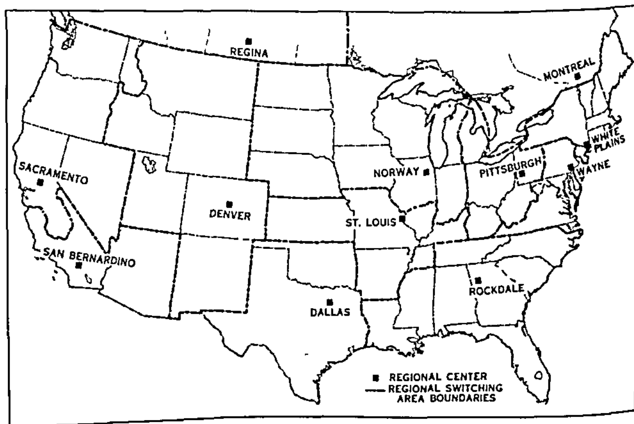


Fig. 3. Regional switching areas of United States and Canada *(American Telephone and Telegraph Co.)*

matic alternate routing, the testing of ... more routes in a predetermined order to find an intercity trunk not in use; (2) selection of the desired route from the first three or first six digits dialed, and (3) code conversion, erasure of digits dialed, and ... of digits to those dialed, often

... primary center handling through-switched traffic may also act as a toll center to handle the originating and terminating traffic for the exchange area it serves

General switching plan. Each local and long distance office is connected to an office of higher rank by means of a group of trunks called a final ... All regional centers are directly interconnected by final groups Figure 3 shows the regional ... areas in the United States and Canada.

...ure 4 shows some typical homing arrangements. The home office may be of any higher rank. ... long-distance offices are the home for end ..., even though this is not shown on Fig. 4. ... allow direct connection between as many offices ... possible, a high usage intercity trunk group is provided between two offices of any rank when ... the cost per trunk is less than the cost of the ... alternate route (two or more intercity trunks in tandem), and (2) there are enough calls to load ... minimum of two or three trunks. Figure 4 shows ... high-usage intercity trunks might be provided ... offices of differing rank. High-usage trunk groups are engineered to carry only a portion of the total calls made during the busiest hours of the ... To maintain a high grade of service, the remaining or overflow calls are offered to an alternate route. The alternate routing capabilities of dial switching systems automatically select a route in which there is an intercity circuit not in use at the moment. Available routes to the distant point are tested in an order of prearranged preference, the most direct route being the first choice route and the final route being the last choice route Where there is only one route provided, a final route is used.

Figure 5 illustrates the application of alternate routing for a long-distance call from Syracuse, ...

International dialing. Whether by operators or customers, international dialing presents many problems which require cooperative effort for solution, because other countries, particularly in Eu...

... of transmission (speech current) levels, and reconciliation of differences in alphabets and dial designs. A number of countries have resolved these problems for calls between or among themselves. Continued progress will eventually unite all countries in a dialing communications network. [C.M.M.]

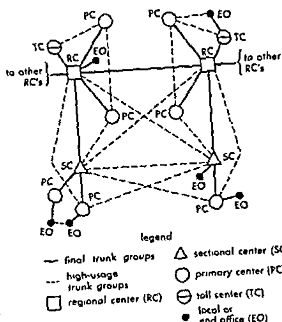


Fig 4 Homing arrangements and high-usage trunks. (American Telephone and Telegraph Co)

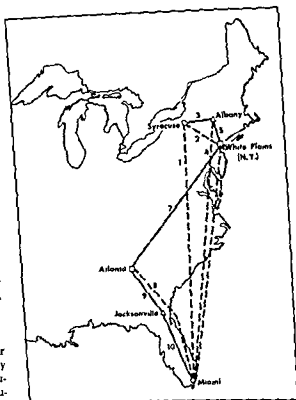


Fig 5 Automatic alternate routing for a long-distance call from Syracuse, N.Y., to Miami, Fla. (American Telephone and Telegraph Co.)

OVERSEAS TELEPHONE SERVICE

Overseas telephone service links the telephone systems of about 140 countries and territories throughout the world. In 1959, there was a total of about 4,000,000 international telephone conversations, 2,700,000 of which involved the United States. Some 96% of all the telephones in the world may be connected with the United States telephone system.

Overseas telephone circuits are largely provided by three transmission systems: high-frequency radio, tropospheric scatter radio, and submarine cable. Each has a definite field of use determined by its own characteristics and limitations.

For example, high-frequency radio systems, the media of overseas telephone communication since the 1930s, have been universally employed because they can link many countries at great distances economically and because their flexibility makes them a practical way to provide a few circuits over long distances. The more recently developed tropospheric scatter radio is primarily useful for communication over inaccessible terrain, such as the Arctic, and relatively short distances over water. Submarine telephone cable systems are not subject to the fading and interference that plague high-frequency radio systems. Cable systems provide superior transmission performance, have great circuit capacity, and can be extended over long distances Although they cost more than high-frequency radio systems, these values have caused them to be used on long water routes

High-frequency radio systems. These systems operate in the 2- to 30-megacycle (Mc) radio-frequency range, provide up to four telephone channels per system, and operate over long distances. For example, the circuit between Oakland, Calif., and Bandoeng, Indonesia, extends 8650 miles. They provide good transmission most of the time but are subject to occasional fading and interference due to disturbances in the ionosphere. The total assigned bandwidth limits the number of such systems that can be used

With only a few exceptions, overseas radiotelephone circuits are obtained through use of single-sideband (SSB) equipment. With SSB, as many as four telephone message circuits can be handled by one radio transmitter. These transmitters have output powers ranging from 2000 to 80,000 watts peak envelope power and are assigned to the various circuits as required by the length of the radio path and transmission conditions encountered on it. For example, circuits from New York to northern European countries, which traverse the auroral zone, require considerably higher power for the same grade of service than circuits from New York to South American countries.

Transmission at these frequencies is accomplished by reflection of the radio waves by the ionosphere and by the earth. The waves may be reflected one or more times before reaching the distant receiver (*see* IONOSPHERE; RADIO-WAVE PROPAGATION) Because of the constantly changing density and position of the ionosphere, the radio frequency is selected for the existing conditions As many as five frequencies in the 2- to 30-Mc range may be required to provide continuous service on some of the longer radio systems during the sun-spot cycle of approximately 11 years.

The specific frequency utilized depends on the time of day and time of year. The highest frequen

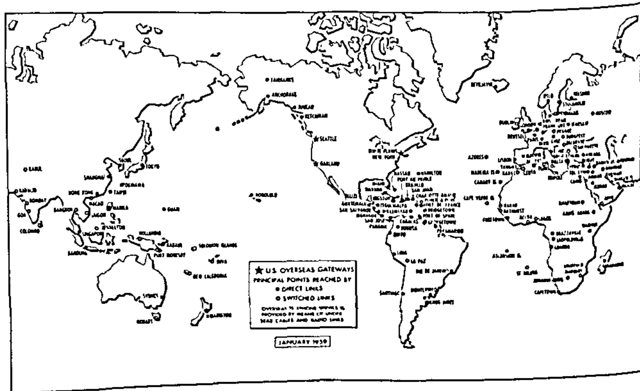


Fig. 6. Bell System overseas telephone service.

s are used during winter daytime, and the low- frequencies are used during winter nighttime.

eration than during the winter months. Radio transmission is also affected by unusual sun-spot tivity. All these factors determine the particular equencies selected to provide good transmission a particular point of reception.

Overseas, the voice currents are restored to oper strength at the radiotelephone receiver. A gnal starting out at 80 000 watts peak envelope ower will often arrive at the input of the distant ceiver as only a few millionths of a watt. At the ceiver this weak signal is amplified and transformed from the radio-frequency range to the oice-frequency range and, by means of filters in e output circuit, separated into the several message channels. Finally, the voice currents are sent ver the domestic telephone system of the distant ountry

A directional antenna is used at the transmitter to focus the radio signals sharply in the desired irection. A directional receiving antenna is used to exclude undesired signals and noise from other directions

Tropospheric scatter systems. These systems normally operate in the 300–3000 Mc frequency range and provide about 60–100 telephone circuits. Although radio waves at these frequencies are not reflected by the ionosphere, it has been found that if large amounts of power are radiated, scattered energy will be received over relatively long distances beyond the horizon. By the use of high-powered transmitters (up to 10 kw) and large, high-

to frequent fading of rather large magnitude, and therefore space-diversity reception is usually re-

to routes over water or inaccessible terrain Radio relay systems are usually more suitable than the scatter systems for overland use where intermediate radio relay stations can be constructed *See* RADIO

Overseas submarine telephone cables. Submarine cables generally consist of a single insulated copper wire and a grounded coaxial return conductor Initially, no underwater amplifiers existed, and the associated telephone equipment was established at the terminals on land. This was satisfactory for transmission over short distances, and such circuits are still used between Florida and Cuba, between Great Britain and the European continent, and many other places.

However, submarine cables over long distances require underwater amplifiers to restore the transmission losses in the cable. These amplifiers, called repeaters, must be capable of operation at 2 or 3 miles below the surface and must have a long life.

Repeaters. There are two basic types of repeaters: the one-way flexible type and the two-way rigid type. The flexible repeater is used in a two-cable submarine system in which each cable operates in a single direction. The amplifier has a flexible housing which passes around the drums and sheaves of the laying gear of the cable ship as part of the cable. The rigid amplifier provides two-way operation in a submarine cable system using only one cable. The amplifier has a rigid steel casing and requires special laying techniques.

In 1956, flexible repeaters were employed in the Newfoundland-Scotland portion of the first transatlantic telephone cable. This type of repeater is spaced at intervals of about 40 miles and is expected to operate for more than 20 years without attention It provides 36 channels at 4-kc spacing.

The repeater is a three-stage feedback amplifier of conventional design. The electron tubes are pentodes of special design to ensure long life. The amplifier provides 60 db gain at the highest frequency. The plate and heater power for the amplifier is supplied over the cable by a constant direct current of 225 milliamps. The plate voltage of approximately 52 volts is derived from the voltage drop across the three heater elements. Built-in testing features are provided whereby each amplifier in the system may be checked by a shore station.

The rigid repeater is usually spaced at 25-mile intervals and provides about 60 channels. These repeaters use a three-stage amplifier that can amplify both directions of transmission. Actually, *two* amplifiers are included. These are connected in parallel with a common feedback circuit. With this *design, one amplifier can fail without seriously affecting* the over-all gain.

Cable. The cable consists of a single coaxial conductor properly armored and protected to withstand strain involved in laying; the pressure and hazards of ice floes, ships' anchors, trawler fishing, and tides, and the pressure of deep water. Through the center of the cable runs a single copper conductor about ⅛ in thick, wrapped with three copper tapes, insulated and separated from the outer conductor by a layer of polyethylene 0 6 in. in out-

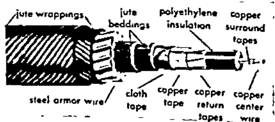


Fig. 7 Submarine telephone cable. (*American Telephone and Telegraph Co*)

side diameter. To complete the electrical transmission path, a grounded return conductor consisting of six thin copper tapes is laid over the polyethylene core.

The use of repeaters in submarine cables has made possible a large increase in overseas facilities for voice transmission over the major routes. However, this availability of circuits and improved transmission has greatly stimulated usage, and consequently methods are being devised to increase the capacity of these submarine cable systems. Thus a new system, called time-assignment speech interpolation (TASI), almost doubles the number of conversations possible on each circuit by utilizing idle time in the unused direction, as well as the pauses between words and even syllables.

New telephone cables and repeaters under development are expected to increase message capacity materially and to permit greater cable length between terminal points. With these improvements, it is believed that such systems will be extended within a few years to all major continents. [C.C.DU.]

TELEPHONE INFORMATION SERVICE

By calling a designated number, current information on telephone numbers and number changes may be obtained. Other services include information on a variety of other subjects.

Telephone numbers. Number information service is generally maintained at a centralized location for a specific area served and is furnished to calling customers by operators. Daily cumulative lists of telephone number changes are usually maintained at these locations to enable customers to call any telephone number not listed in their current number-service directories Number-service information bureaus also function importantly as a source of telephone numbers for the completion of long-distance calls and similar situations where that information is normally not readily available to the caller.

Time. In many areas customers can secure the time of day by calling a designated telephone number. Time-announcement calls are routed to an automatic time-announcement machine or operators located in or near the central-office switching equipment at a centralized location. Time service may be sponsored and provided by the telephone company, or the telephone company may provide facilities for customer-sponsored time services

Weather. Customers may also be able to obtain weather bureau forecasts by calling a designated telephone number. The weather announcements are usually made by mechanical announcement equipment at a centralized bureau. At the centralized bureau regular weather forecasts are received at periodic intervals, usually by teletypewriter from the local office of the weather bureau service. Special bulletins may be issued from time to time depending upon weather conditions. Forecasts are read by operating personnel into a transmitter, and a record of the operator's voice is electrically impressed on a suitable recording media. The record is then connected to a bulletin circuit associated with the designated telephone number. The information is repeated continuously until it is replaced by a new bulletin.

Miscellaneous announcements. Other announcement services, such as news items, train or plane schedules, theatre program information, religious messages, and stock-market quotations, may be made available by arranging to call a designated telephone number which terminates on recorded announcement facilities. This service is usually provided and sponsored by commercial customers of the telephone company. The demand for these services must be carefully administered; otherwise it would be necessary to provide excessive amounts of facilities to avoid overloads which may affect the quality of other telephone services. [C.M.M.]

TELEPHONE LEAVE-WORD SYSTEMS

Special services are often available to permit callers of a particular line to leave messages in the absence of the called subscriber. These are of two general types: secretarial systems and telephone answering sets.

Secretarial systems. These systems employ switchboards similar to manual PBXs but have features designed especially for secretarial service.

regular telephone line at the central office serving the latter See TELEPHONE PRIVATE BRANCH EXCHANGE (PBX).

The attendants at the switchboard answer calls if they are not answered by the subscriber after a prearranged number of ringing cycles. The calling customer can then leave any desired message with the attendant When the subscriber to the secretarial service returns to his office or home, he calls the secretarial switchboard, and any messages received during his absence are read to him. These services are usually provided by firms organized for the purpose, using switchboards and wire facilities provided by the telephone companies.

Because of the distances involved where a given secretarial bureau may serve subscribers in outlying areas of the larger cities, the cost of providing direct-wire facilities to each subscriber may present an economic problem For this reason the concentrator-identifier principle is often utilized to obtain more efficient use of the conductors. An automatic switching apparatus, called a concentrator, is located at a central office in the area being served. All calls to subscribers in that area are identified on other equipment at the secretarial service by causing the proper lamp to light. A relatively small number of conductors are required for connecting the two devices There will be, of course, moments of temporary overload when the number of calls exceed the facilities, but these are no greater than those for telephone service generally.

. . . T' 'omatic prem- ...er in-
ng calls with a recorded message in his nce and, in one type, permit the caller to re- a message in turn. Only the latter type comes in the scope of leave-word services.

...ese machines generally use magnetic record- and reproducing techniques, although some use ...ooved disk-type recording for the answering ...sage The magnetic medium most widely used band of synthetic rubber, impregnated with oxide particles, fitted tightly on a drum of magnetic metal. Some machines use steel wire ...he recording medium [R.F.D.]

PRIVATE-LINE TELEPHONE SERVICE

...n its most elementary form private-line service ...sists of two telephone instruments, or stations, ...manently connected to each other by a suitable ...smission path, and a means of signaling so that ...er station can indicate to the other when a con- ...sation is to be transmitted. This service is not ...ited to the connection of two points only but ...y also include multipoint networks and more ...plex signaling arrangements. In the United ...tes at the end of 1958 there were approximately ...00,000 circuit miles of line facilities under lease ... private-line telephone service.

For many applications, private-line service offers ...nomic and technical advantages. On the eco- ...mic side, private-line service will generally re- ...lt in significant monetary savings for the cus- ...mer if his telephone traffic is heavy enough and ...llows a fairly constant geographic pattern. The ...chnical advantages lie in the capability of pri- ...te lines to meet specialized complex require- ...ents, principally those of large users. The provi-

...ouping arrangements, which can at present only ...e furnished on a private-line basis. The reason for ...is is that private-line service always involves the ...se of communication channels with predetermined ...erformance requirements. Consequently, when ...ecessary, the line facilities and stations compris- ...ng it can be engineered to meet more rigid techni- ...al requirements than would otherwise be the case

Of particular importance in this regard is the ...mpedance match obtainable between telephone in- ...struments and lines, and between line sections ...themselves. Since the line-facility layout of a pri- ...vate line system is fixed, better impedance matches can generally be made, resulting in a higher grade of transmission than that which would be obtained by connecting the same points through the regular switched message network.

The term "private line" has often been incor- ...rectly used to describe a grade of telephone service in which a subscriber is connected to his serving exchange office via a line assigned for his use only. Service of this nature is described by several terms, the most prevalent of which are individual line or one-party service. [R.R.HO.]

TELEPHONE TRANSMISSION RATING

A routine procedure is followed by the Bell System operating companies to measure and grade the over-all quality of telephone transmission. In effect it is a plan for rating the relative ease with which *telephone conversations may take place after* the connection has been established. Many factors affect this, including accuracy of reproduction of speech and freedom from extraneous noises. In the modern telephone system these effects have been reduced so much that a good comparative measure of telephone transmission quality, from the customer's point of view, is the volume, or level, of speech received from the distant talker. Speech volume measurements, therefore, are used as the basis for rating transmission.

Received volumes are measured most accurately at the customer's telephone set. It would be difficult and expensive, however, to obtain large amounts of data in this fashion. Therefore, the measurements are taken at the central office, and an average value is assumed for the attenuation of speech volumes which occurs in the connections between the subscribers and the central office.

To measure electrical speech volumes a meter known as a volume indicator is connected successively to a large number of circuits over which local calls are being placed. The volume indicator is essentially a sensitive ac voltmeter with a frequency characteristic and speed of response which permit the needle to deflect in proportion to the energy contained in the electrical speech wave at any instant. A typical speech-volume pattern is shown in Fig. 8. The measurement of peak amplitudes during a given interval of time bears a statistical relationship to the actual loudness of speech being received by the telephone user. These peak amplitudes, then, are used to obtain the volume rating.

The scale of the volume indicator is divided into sectors corresponding to "good," "fair" and "poor," the dimensions of these sectors having been determined by comprehensive tests of customer reaction to a wide range of received volume levels When the volume peaks of the distant party

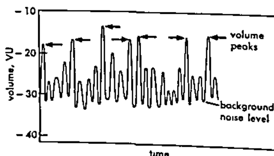


Fig 8 Typical speech volume pattern.

(assumed to be the lower volume of the two parties) have caused the needle to deflect into a particular sector several times within a specified period of time, the call is graded accordingly.

For convenience in the detection and analysis of substandard transmission conditions, the volume data on local calls is separated and classified as intraoffice, between customers served by the same central office; direct-trunked, served by two different offices linked by direct-trunk transmission facilities; and tandem-trunked, served by two different offices linked by trunks which pass through an intermediate switching point. The transmission on each of these may be affected by different factors. A sample of 200 calls of each classification has been found to be adequate for comparison. Therefore, an office rating is presently based upon a minimum sample of 400 or 600 measurements, depending upon whether or not tandem trunk routes are used.

The data are summarized in terms of the percentage of "good," "fair" and "poor" calls in each category. The average number of intraoffice, direct-trunked, and tandem-trunked calls per day is then used to compute a weighted result for the office. The weighted office result is used to enter a standard chart from which the office rating is obtained. [H E CO.]

Bibliography: Bell Telephone Lab., *Transatlantic Submarine Telephone Cable System*, Monograph 2710, 1957, K. Bullington, Characteristics of beyond-the-horizon radio transmission, *Proc IRE*, 43(10):1175–1180, 1955; C. C. Duncan, The transatlantic telephone cable puts out to sea, *Bell Telephone Magazine*, 34:93–95, 1955; R. A. Felsenheld, et al., Wide-band ultra high-frequency over-the-horizon equipment, *Electrical Engineering*, 35:86–93, 1958; T. J Morgan, *Telecommunication Economics*, 1958, J. R. Pierce, *Electrons, Waves and Messages*, 1956, J Poole, *The Telephone Handbook*, 1946; W. G Radley, *Fifty Years' Development in Telephone and Telegraph Transmission in Relation to the Work of Heaviside* 1950; W. H Tidd, Demonstration of bandwidth capabilities of beyond-horizon tropospheric radio propagation, *Proc. IRE*, 43(10):1297–1299 1955

Telephone signaling, special subscriber

A system supplementing or replacing the usual telephone bell, to increase the area covered or otherwise provide more comprehensive means for notifying subscribers that they are wanted at the telephone.

Functionally, these are of two types, those which give extended coverage without indicating which individual is wanted, and those which indicate that a particular individual is wanted.

The first type of system uses audible or visual signaling systems, generally operated in synchronism with the ringing cycle of the central office or private branch exchange (PBX) ringing source.

Audible signals. These include a simple extension ringer (similar to the telephone ringer), musi-

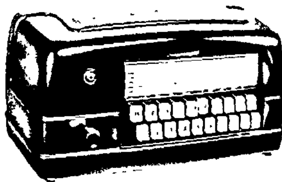


Fig. 1. Manual code calling system control.

cal gongs, loud-ringing bells, and horns. Those which require low power may operate directly from the ringing source. Those requiring greater power use a relay, activated by the ringing source, to connect commercial power to the signaling device.

Visual signals. These utilize neon lamps, which light directly on ringing current, or incandescent lamps which light from a local power source under control of a relay activated by the ringing current. Relay sets can be connected to any type of light, such as desk lamps, bed lamps, and flood lights, which operate on commercial power.

Individual paging systems. Special systems which give an indication that a particular individ-

associated with PBXs. They sound a prearranged code number throughout the premises, usually by signals on single-stroke musical gongs. With manual PBXs, the PBX attendant presses the button (see Fig 1) numbered with the desired individual's code, and the mechanism sends out that code, repeating it a fixed number of times or until the attendant stops it. Code calling systems with dial PBXs may consist of switching and relay equipment whereby any extension user can dial the code of the individual wanted. In a typical system of this type three-digit codes, using digits from one to five, are employed. This gives a maximum of 125 codes.

Loudspeaker paging systems. These systems are often associated with PBXs to call individuals by name. They may use a special microphone or the transmitter of the PBX attendant's regular telephone headset, transferred temporarily to the loudspeaker system by operating a key.

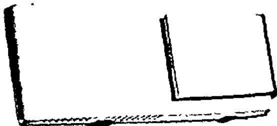


Fig 2 Radio paging recei

Radio paging systems. These comprise miniature dio receivers, carried in a pocket. The receiver ves the wearer an audible signal in response to a dio-frequency signal actuated by a telephone erator at a nearby long-distance switchboard. gure 2 shows a typical pocket receiver used in ese systems. See TELEPHONE PRIVATE BRANCH CHANGE (PBX); TELEPHONE SERVICE. [R F.D.]

elephone systems construction

he process of physically constructing the outside lant portion of a complete telephone system rincipal components consist of (1) subscriber istribution facilities from the central office to the ustomer's premises, (2) interoffice trunks, and (3) interurban cable or wire facilities. This construction includes new construction and reinforcement or replacement of existing facilities. The construction process includes the pole lines and conduit systems, as well as the wires and cables. It also includes joining and testing the circuits to insure electrical integrity. Circuits are provided over cable or wire.

Cables and wires. Cables consist of individually

Cables are identified according to their use—aerial, underground, or buried.

Wire circuits are open wire or multiple wire, except for a small amount of buried wire. Open wires are individual, bare, spaced wires which are supported by crossarms or bracket-mounted insulators on poles. Multiple wire consists of insulated conductors that are stranded around an sulated high-strength steel support wire but does ot have the outer sheath of a cable. Multiple wire as from 6 to 26 pairs. The support wire is usually 109 or 0 120 in in diameter, depending on strength equirements.

or economic reasons, communication lines and ower lines often share the same poles. This is lso done to avoid the objections to two adjacent arallel pole lines.

Mechanized equipment is generally used to dig the pole hole and set the pole. A two-man crew with line truck can set all but the larger joint poles. Figure 1 shows a modern line truck being used to set a pole.

Where there is an appreciable bend in a pole line, an anchor and guy are placed to balance the side load. Mechanized anchors are attached to anchor rods and inserted in holes dug in the ground. The anchors are then expanded to engage undisturbed soil. When the anchor is secure, the hole is backfilled and tamped and a guy is placed from the anchor rod to the pole.

If open wire is to be placed on the pole line, crossarms are attached to the poles before they are set.

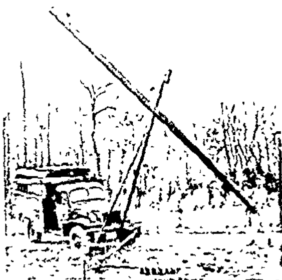


Fig 1 Two-man line truck with front-mounted derrick being used to set pole. The previously used digging attachment is on the ground in front of the truck.

Aerial wire. Although there is still a substantial amount of open wire in service, little has been placed since World War II and that in use is being rapidly replaced. Open wire has a high cost per circuit and is vulnerable to storms.

Multiple line wire, commonly known as urban and rural wire, was introduced around 1954. It permits economical and rapid placement, although it has a physical life somewhat shorter than conventional cable Figure 2 shows such wire at a pole attachment

Aerial cable Aerial cable is generally lashed to a supporting steel strand which is attached to the poles. Where the volume of work justifies the equipment, the prelashing method of placing the cable is the most economical. By this method, the strand

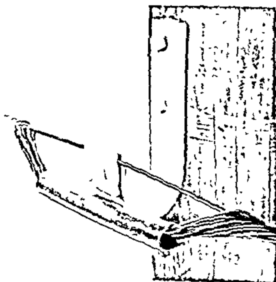


Fig 2 Multiple line wire at pole support Large wire in clamp is insulated, high-strength, steel support wire.

(assumed to be the lower volume of the two parties) have caused the needle to deflect into a particular sector several times within a specified period of time, the call is graded accordingly.

For convenience in the detection and analysis of substandard transmission conditions, the volume data on local calls is separated and classified as intraoffice, between customers served by the same central office; direct-trunked, served by two different offices linked by direct-trunk transmission facilities; and tandem-trunked, served by two different offices linked by trunks which pass through an intermediate switching point. The transmission on each of these may be affected by different factors. A sample of 200 calls of each classification has been found to be adequate for comparison. Therefore, an office rating is presently based upon a minimum sample of 400 or 600 measurements, depending upon whether or not tandem trunk routes are used.

The data are summarized in terms of the percentage of "good," "fair" and "poor" calls in each category. The average number of intraoffice, direct-trunked, and tandem-trunked calls per day is then used to compute a weighted result for the office The weighted office result is used to enter a standard chart from which the office rating is obtained [H.E.CO]

Bibliography: Bell Telephone Lab., *Transatlantic Submarine Telephone Cable System*, Monograph 2710, 1957; K. Bullington, Characteristics of beyond-the-horizon radio transmission, *Proc. IRE*, 43(10):1175–1180, 1955; C. C Duncan, The transatlantic telephone cable puts out to sea, *Bell Telephone Magazine*, 34:93–95, 1955, R. A. Felsenheld, et al., Wide-band ultra high-frequency over-the-horizon equipment, *Electrical Engineering*, 35:86–93, 1958; T. J. Morgan, *Telecommunication Economics*, 1958; J R Pierce, *Electrons, Waves and Messages*, 1956, J. Poole, *The Telephone Handbook*, 1946; W G. Radley, *Fifty Years' Development in Telephone and Telegraph Transmission in Relation to the Work of Heaviside*, 1950; W. H. Tidd, Demonstration of bandwidth capabilities of beyond-horizon tropospheric radio propagation, *Proc. IRE*, 43(10)·1297–1299 1955

Telephone signaling, special subscriber

A system supplementing or replacing the usual telephone bell, to increase the area covered or otherwise provide more comprehensive means for notifying subscribers that they are wanted at the telephone.

Functionally, these are of two types: those which give extended coverage without indicating which individual is wanted, and those which indicate that a particular individual is wanted.

The first type of system uses audible or visual signaling systems, generally operated in synchronism with the ringing cycle of the central office or private branch exchange (PBX) ringing source.

Audible signals. These include a simple extension ringer (similar to the telephone ringer), musical gongs, loud-ringing bells, and horns. Those which require low power may operate directly from the ringing source. Those requiring greater power use a relay, activated by the ringing source, to connect commercial power to the signaling device.

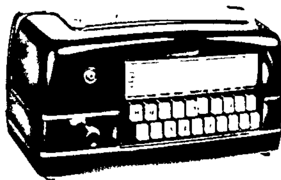


Fig. 1. Manual code calling system control.

Visual signals. These utilize neon lamps, which light directly on ringing current, or incandescent lamps which light from a local power source under control of a relay activated by the ringing current. Relay sets can be connected to any type of light, such as desk lamps, bed lamps, and flood lights, which operate on commercial power.

Individual paging systems. Special systems which give an indication that a particular individual is wanted include code calling systems, loudspeaker paging systems, and radio paging systems.

Code calling systems. These systems are usually associated with PBXs. They sound a prearranged code number throughout the premises, usually by signals on single-stroke musical gongs. With manual PBXs, the PBX attendant presses the button (see Fig. 1) numbered with the desired individual's code, and the mechanism sends out that code, repeating it a fixed number of times or until the attendant stops it Code calling systems with dial PBXs may consist of switching and relay equipment whereby any extension user can dial the code of the individual wanted In a typical system of this type three-digit codes, using digits from one to five, are employed. This gives a maximum of 125 codes.

Loudspeaker paging systems These systems are often associated with PBXs to call individuals by name They may use a special microphone or the transmitter of the PBX attendant's regular telephone headset, transferred temporarily to the loudspeaker system by operating a key

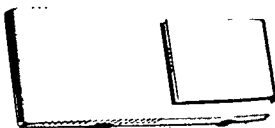


Fig 2 Radio paging receiver.

..io paging systems. These comprise miniature ... receivers, carried in a pocket. The receiver ... the wearer an audible signal in response to a ...-frequency signal actuated by a telephone ...ator at a nearby long-distance switchboard ...re 2 shows a typical pocket receiver used in ... systems. *See* TELEPHONE PRIVATE BRANCH ...ANGE (PBX); TELEPHONE SERVICE. [R F D.]

...phone systems construction

...process of physically constructing the outside ...t portion of a complete telephone system. ...cipal components consist of (1) subscriber ...ribution facilities from the central office to the ...omer's premises, (2) interoffice trunks, and ...interurban cable or wire facilities. This con...ction includes new construction and reinforce...t or replacement of existing facilities The ...struction process includes the pole lines and ...duit systems, as well as the wires and cables. It ...includes joining and testing the circuits to in...e electrical integrity. Circuits are provided over ...le or wire.

...ables and wires. Cables consist of individually ...ulated conductors that are stranded into a core

...derground, or buried.

Wire circuits are open wire or multiple wire, ...cept for a small amount of buried wire Open ...ires are individual, bare, spaced wires which are ...pported by crossarms or bracket-mounted in...lators on poles Multiple wire consists of insu...ted conductors that are stranded around an

...equirements.

Pole lines. Pole line design is based upon the ...erial structures to be supported and the average ...inter weather conditions typical for the location. ...or economic reasons, communication lines and ...ower lines often share the same poles. This is ...lso done to avoid the objections to two adjacent parallel pole lines

Mechanized equipment is generally used to dig the pole hole and set the pole. A two-man crew with line truck can set all but the larger joint poles Figure 1 shows a modern line truck being used to set a pole.

Where there is an appreciable bend in a pole line, an anchor and guy are placed to balance the side load Mechanized anchors are attached to anchor rods and inserted in holes dug in the ground. The anchors are then expanded to engage undisturbed soil When the anchor is secure, the hole is backfilled and tamped and a guy is placed from the anchor rod to the pole

If open wire is to be placed on the pole line, crossarms are attached to the poles before they are set

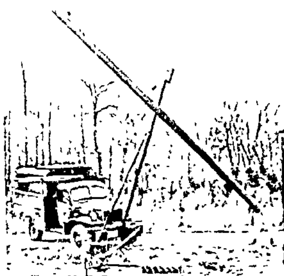


Fig 1 Two-man line truck with front-mounted derrick being used to set pole. The previously used digging attachment is on the ground in front of the truck.

Aerial wire. Although there is still a substantial amount of open wire in service, little has been placed since World War II and that in use is being rapidly replaced Open wire has a high cost per circuit and is vulnerable to storms

Multiple line wire, commonly known as urban and rural wire, was introduced around 1954 It permits economical and rapid placement, although it has a physical life somewhat shorter than conventional cable Figure 2 shows such wire at a pole attachment

Aerial cable. Aerial cable is generally lashed to a supporting steel strand which is attached to the poles. Where the volume of work justifies the equipment. the prelashing method of placing the cable is the most economical By this method, the strand

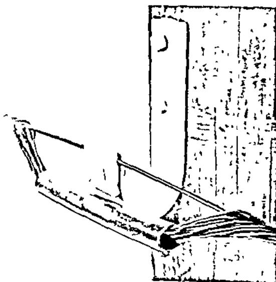


Fig 2 Multiple line wire at pole support. Large wire in clamp is insulated, high-strength, steel support wire



Fig. 3. Six-duct clay conduit being joined by mortar bandages.

and cable are properly tensioned and lashed together as they are pulled into place. If the prelashing technique is not used, the steel strand is placed and tensioned first. The cable is then lashed to the strand as it is being placed.

Terminals Terminals must be provided on subscriber distribution cables to permit attachment of service wires to the customer's premises Until about 1954, the paper and pulp insulation in such cables necessitated hermetically sealed terminals to prevent entrance of moisture. Only those cable pairs connected to the terminal binding posts are available for use at such terminals Plastic was introduced as conductor insulation in subscriber distribution cables in 1954. This insulation is impervious to atmospheric moisture, and it is therefore possible to use ready-access terminals, which give access to every conductor in the cable.

Underground lines. Underground lines are placed in conduit, which runs between manholes placed at 400- to 1000-ft intervals. A truck winch line is used to pull cable from a reel mounted above one manhole through the conduit to the next manhole, or through several intervening manholes if branching is not required and the length is not too great. Manholes are usually of concrete cast in position; however, they may also be precast.

Conduit is available in a number of materials including concrete, fiber, iron pipe, and clay. Multiple clay duct with 3, 4, 6, 8 or 9 ducts is the predominant type. Adjoining pieces of conduit are held in alignment by dowel pins inserted in matching holes in the conduit. Mortar bandages are wrapped around the junction, resulting in a strong, watertight bond (see Fig. 3).

Underground cable may contain as many as 2100 pairs of conductors. A full reel of 2100-pair, 26-gauge, plastic sheath cable is about 1350 ft long and weighs approximately 3 tons.

Where the number of cables along a given run are not sufficient to justify conduit, the cable may be buried directly in the ground. In some situations the cable may be buried with no outer protection. However, if rodents, corrosion or physical hazards require, the cable may be protected with jute or various types of armor. Buried interurban cables are usually plowed in with heavy plow equipment at depths of 3–5 ft (see Fig. 4). Buried urban or rural cables are usually placed in a trench or plowed with lighter equipment at depths of from 2 to 3 ft.

Cable splicing. After cables are placed, they must be spliced to provide a complete system. The individual conductors must be joined and the sheath opening closed Before 1958, conductor joining was accomplished by skinning off the insulation, twisting the skinned conductors to form a pigtail joint and slipping an insulating cotton sleeve over the joint (see Fig 5).

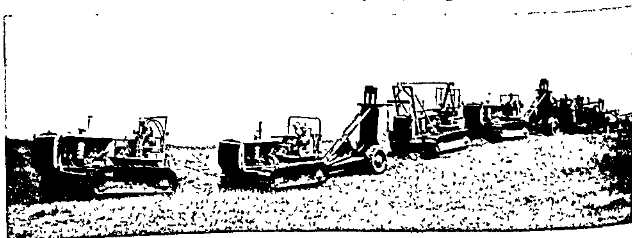


Fig. 4. Typical, heavy plow train placing intercity buried cable. First plow is a router; the second places the cable.

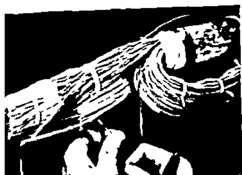


Fig. 5 Pigtail splice between sections of a pulp-insulated cable.



Fig. 6. Half-section of a splice case, ready for closing. Identical section is bolted to that shown. Note plastic cords along sides of case and plastic collars at ends to insure seal.

The present conductor-joining technique is known as the punch sleeve method. This technique applies only to paper and pulp insulated conductors. The unskinned conductors to be joined are inserted into a plastic insulated copper tube. The tube or sleeve is then pressed between the jaws of a ...

made pigtail joint.

After all conductors are joined, the sheath opening is closed. A mechanical splice case is generally used for this purpose (see Fig. 6). [G L C]

Telephony

The transmitting of speech to a distant point by means of electric waves. There are three elements basic to telephonic communication: (1) telephone sets to change sound energy into electric waves and back again (see TELEPHONE); (2) transmission systems to carry electric waves over any distance within acceptable limits of distortion or attenuation (see TELEPHONE SERVICE); and (3) switching systems to connect any two telephone sets, see SWITCHING SYSTEMS (COMMUNICATIONS).

Beginnings and growth. The telephone was invented by Alexander Graham Bell and patented in the United States in 1876. Commercial telephone service was inaugurated by renting pairs of telephone sets to individuals for local service. A single iron wire with a ground return provided the connection. Speech transmission was poor and was limited to a distance of a few miles.

Figure 1 illustrates this elementary telephone circuit. One device serves as both transmitter and receiver. Sound waves striking the diaphragm cause it to vibrate. The motion of the diaphragm changes the magnetic field, inducing electric waves of varying voltage and current in the winding. These travel to the distant telephone, where the changes produced in the magnetic field cause the diaphragm to reproduce the original sound. Later development provided separate transmitters and receivers with auxiliary circuitry to bring the reproduction of speech up to high standards of fidelity.

The need to connect any two of a large number of telephone sets led to the development of a switchboard in 1878. The advantage of a central switching office is illustrated in Fig. 2 for a simple case of five modern telephones. When directly connected, 10 lines are needed to enable any two to be interconnected. In addition there is the problem of providing either four telephones or some switching arrangement at each customer's location. Only five lines are needed with a central switching office, and a switchboard can take the place of individual switches to interconnect any two lines. With a switchboard, each line is associated with a socket called a jack and any two lines are interconnected by the operator with plug-ended cords inserted into the proper jacks. Other switchboard circuitry enables the operator to talk, listen, and signal as necessary.

Efforts to interconnect any two telephone sets automatically were successful in 1889 with the invention of a practicable mechanism by Almon B. Strowger. Thus a verbal request for a connection to a distant telephone was replaced by dial signals initiated by the customer. Subsequent research and development improved the step-by-step switch and led to other systems of greater complexity. This complexity took the form of memory and logic functions analogous to computer systems. As the

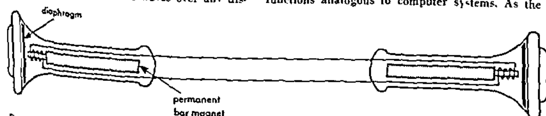
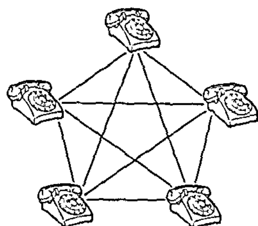
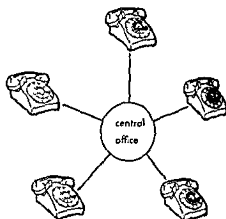


Fig. 1. Elementary telephone circuit (American Telephone and Telegraph Co.)



(a)



(b)

Fig. 2 Two ways of interconnecting five telephones. *(a) All direct. (b) All switched (American Telephone and Telegraph Co.)*

number of telephones to be interconnected grew, larger cities needed several and then many central offices. The interconnection problem shown in Fig 2 was then of a higher order of magnitude, central offices taking the place of the telephone sets shown. Central offices are interconnected by interoffice trunks, sometimes directly. In other cases, particularly in the larger cities, a central switching location called a tandem office can greatly reduce the total number of interoffice trunks needed For interurban service, the interconnection problem presents a still higher order of magnitude. To connect any two of millions of telephones automatically, memory is needed to retain the number called and some of the actions taken; logic is needed to decide the proper action to take from a number of possible alternatives which vary from call to call Central offices are connected to long-distance offices by connecting trunks. Long-distance offices are interconnected by intercity trunks Dial switching systems are now used in most of the countries of the world for local service and in many countries to a varying extent for interurban service. The table illustrates the widespread use and automation of telephony.

International telephony. Both land lines and radio relay systems are employed in international systems. Water barriers are bridged by undersea cables or radiotelephone systems. Installation and operation of equipment and systems used for international telephone service take place by agreement between governmental agencies or private companies, as the case may be for each country. To promote international cooperation, exchange of information, and establishment of recommended practices, the International Telecommunication Union (ITU), now affiliated with the United Nations, was founded in 1865. Two organs of the ITU act as consulting groups in technical matters, the International Telephone and Telegraph Advisory Committee (CCITT) and the International Radio Advisory Committee (CCIR)

Countries with 1,000,000 or more telephones

Country	Number of telephones (Jan. 1, 1958)	Telephones per 100 population	Automatic operation, %	Connecting with U S telephones, %	Conversations per capita for the Year 1957
Argentina	1,181,121	5 89	83.0	99.9	168 2
Australia	1,873,791	19.22	71.0	100	138 9
Canada	4,816,118	28 64	79 9	99 8	497 1
France	3,498,900	7.92	72.5	100	†
Germany, Dem. Rep.	1,123,278	6 45	92 5	100	52 8
Germany, Fed. Rep	4,731,945	8.75	97 7	100	66 5
Italy	2,871,011	5.91	96.0	97.7	98 9
Japan*	3,886,327	4.22	60 6	100	118 0
Netherlands	1,318,269	11.88	96.9	100	111 8
Spain	1,339,653	4.51	78.6	100	90.0
Sweden	2,409,842	32 60	82 0	100	190.5
Switzerland	1,385,125	26 84	99.7	100	198 6
U.S S R	3,558,000	1.78	46.1	60	†
United Kingdom*	7,354,000	14.25	79 2	100	78.5
United States	63,620,900	36 82	89.4	100	460 0
World	117,800,000	4 2	83 9	97.0	†

* March 31, 1958. † Not available.

Telephony in the United States. The American Telephone and Telegraph Company, 20 operating subsidiaries, and two associated companies comprise the Bell System and serve about 85% of the nation's telephones. Several thousand independent companies serve the remaining telephones, virtually all of which can be connected to the Bell System. Approximately 90% of United States telephones have local dial service. Long-distance dial service, called direct distance dialing (DDD), was inaugurated in 1951 and is rapidly becoming as widespread as local dial service. Many industries, businesses, and government agencies lease telephone channels (talking paths) for their exclusive use. This is called private-line telephone service. The transmission systems used for telephony can be used for many other communication services, for examples, see DATA TRANSMISSION; FACSIMILE; RADIO BROADCASTING NETWORKS; TELEGRAPHY; TELEPHOTOGRAPHY; TELETYPEWRITER EXCHANGE (TWX) SERVICE; TELEVISION NETWORKS.

Manufacture of telephone equipment. Most equipment for switching and transmission systems is manufactured on an assembly-line basis. Apparatus, devices, sockets, and brackets are fastened to mounting plates or other supports, which are then assembled on steel frameworks of various

to interconnect the units of equipment that comprise a working system. Advantage is taken of miniaturization, printed wiring, and use of plug-in units to facilitate manufacture, installation, and maintenance. See SYSTEMS ENGINEERING.

Fundamental planning. The telephone communication complex of the United States is the result of an evolutionary technology. New system and equipment designs must work in compatibility with older systems and equipment and must possess flexibility to permit the introduction of still newer technology without any major displacement of existing instrumentalities. Older equipment is often

important that telephone service be economically attractive to both the telephone users and the telephone industry. Short- and long-range planning as well as nationwide and international coordination of all aspects of telephony, receive regular attention to maintain fast, accurate, and reliable service while demands for more service of wider scope continue to grow. See COMMUNICATIONS SYSTEMS (TRAFFIC) DESIGN. [C.M.M.]

Bibliography: American Telephone and Telegraph Co., *Principles of Electricity Applied to Telephone and Telegraph Work*, 1953; C. S. Berkeley, *Traffic and Trunking Principles in Automatic Telephony*, 2d ed., 1949; N. V. Knight and W. Prickett (eds.), *The Telephone Handbook*, 8th ed., 1942; J. R. Pierce, *Electrons, Waves and Messages*, 1956.

Telephoto lens

A photographic lens system specially designed to give a large image of a distant object in a camera of relatively short focal length. A telephoto lens generally consists of a positive lens system and a negative lens system, separated by a considerable distance. If color correction is desired, each of the partial systems must be color-corrected. It is usually not easy to correct distortion in a teleobjective, but occasionally it has been achieved.

Teleobjectives of very long focal length and with infrared correction have become important for aerial reconnaissance, where the reduction of the back focus is necessary because of space considerations. See CAMERA; LENS, OPTICAL. [M.H.]

Telephotography

The transmission of photographs over electrical communication channels. The channels are generally those provided by common carrier communication companies. Individual users furnish telephotograph or facsimile machines to suit their requirements. In the United States such services range from simple arrangements connecting two sets of machines within a city to large networks involving many cities and types of communication facilities. Coordination must thus be effected between the communication companies and the manufacturers of equipment used by customers for their stations to insure compatibility of equipment and optimum system performance. Unattended operation at receiving points is made possible by starting and synchronizing the customer's machines with pulses or tones generated at the transmitting location.

The transmission of news photographs and weather maps is one example of a use with contrasting requirements involving different types of machines and networks. Normally, news pictures require half-tone reproductions with a number of shades of gray. At the receiving locations photographic reproducing devices are used. In contrast, weather maps and other black and white nonshaded material can readily be reproduced by direct recording processes. Networks used for transmitting news pictures are designed for transmitting and receiving at most points because of the unpredictable nature of the location of news events, while the weather network has only a few transmitting points at key locations with many receiving-only locations.

For coordination of operations and for making special announcements, both types of networks must handle speech. Therefore, loudspeakers and telephones are provided at all locations.

Most of the equipment in use today operates at speeds which will transmit an 8½ × 11-in. page in about 6 minutes. There has been considerable interest in recent years in much higher speeds of

transmission, which would require wider bandwidths, and in equipment that will produce sharper definition so that the resulting photographic film may be used for engraving plates suitable for printing of high quality. A number of trials have been run over special intercity facilities in which definitions up to 1000 lines/in. at high speeds have been tested. This service may require bandwidths ranging from 25 kc–1.0 Mc.

Today, the usual transmission equipment employs a double sideband or vestigial single sideband amplitude-modulated carrier for transmission Representative carrier frequencies are 1920, 2000, and 2400 cps; thus voice bandwidth circuits can be used.

Photographs, with their various shades of gray, require the transmission of considerably more information than simple black and white material. Consequently, circuits used to transmit photographic information must be specially engineered and maintained. For example, to obtain sharp edges on picture material, it has been found important to reduce transmission distortion to the extent that the elemental areas of the scanning process are not shortened or elongated by more than one-half of their size. This means that the high-grade intercity networks must be equalized from about 1200 cps to about 2600 cps to within about 300 microseconds of envelope delay. Amplitude-frequency equalization is also important and must be maintained within reasonable limits over the transmitted band.

Special treatment is given facilities used for the permanent networks. Abrupt changes in level, which could cause sharp changes of shading in pictures, are minimized. Reflection, or echo currents, which could cause multiple images, are reduced. Random and impulse noise, which could cause interfering patterns, are avoided. The linearity of amplifying equipment is improved to maintain a wide gray scale.

The transmission of nonshaded black and white material requires the same general techniques, but broader limits are permissible with respect to level variations and certain types of noise. This is particularly true when direct-recording processes are used or where high quality is not a paramount consideration. See FACSIMILE. [C.C DU.]

Telescope

An optical instrument which increases the apparent angular width of distant objects, terrestrial or astronomical, in order to resolve them The simplest type of telescope consists of an objective lens, which forms a real image of the object, and an ocular or eyepiece for magnifying and viewing this image (*see* EYEPIECE; LENS, OPTICAL). A telescope having an objective lens is called a refracting telescope (Fig. 1*a*). The objective lens can be replaced by a mirror or by a system containing refracting and reflecting elements (catadioptric system). Many of the telescopes used for astronomical observations have a concave parabolic mirror, rather than a lens, as the objective. These are called reflecting telescopes (Fig 1*b*)

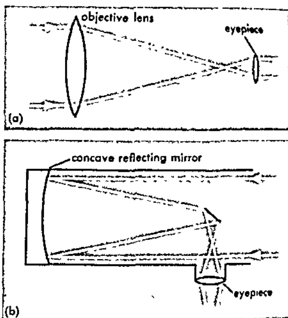


Fig 1. Simplified diagrams of (a) refracting telescope (Kepler's type) and (b) reflecting telescope (Newtonian type).

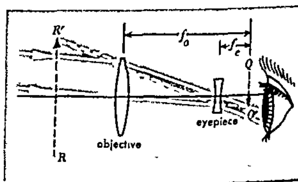


Fig 2. Diagram of a Galilean telescope.

Types. The Galilean telescope, a refracting terrestrial telescope that was first constructed by Galileo in 1609, is shown in Fig. 2 In this instrument [illegible] light alone [illegible]

the Galilean telescope forms erect images, in contrast to the astronomical telescope, which forms inverted images To the observer at the right the light rays appear to come from *RR'*, the enlarged virtual image For a distant object the magnification is f_o/f_e, where f_o and f_e are the focal lengths of the objective and eyepiece, respectively. In addition to the advantage of forming an erect image, the Galilean telescope has the advantage of forming an image which is relatively bright for night observations. However, these advantages are counterbalanced by the fact that its field of view is smaller than that of the astronomical telescope and also by the fact that the image is less sharply defined. In low magnifications, the Galilean telescope appears in common opera-glass constructions. *See* IMAGE, OPTICAL; OPERA GLASS [illegible]

The objective for an astronomical telescope is a lens system of long focal length corrected for aperture, asymmetry, and chromatic errors. For astronomical photography, photographic lenses of large focal length and with especially good corrections are often used. *See* ASTRONOMICAL PHOTOGRAPHY; TELESCOPE, ASTRONOMICAL.

Modern terrestrial telescopes are essentially astronomical telescopes containing mirrors, prisms, or additional lenses to reinvert the image. For details on optical instruments which utilize the principle of the terrestrial telescope, *see* BINOCULARS; GUN SIGHTS; PERISCOPE; RANGEFINDER, OPTICAL. Prisms have the advantage of combining the necessary length of light path with a short length of the instrument

For telescopes of large focal length, and especially for systems which are used outside the visible spectrum, the correction of secondary spectrum becomes difficult. Therefore in many cases reflectors or, more recently, catadioptric systems such as the Schmidt camera, are used, in which the mirror part provides all the necessary power while the catadioptric auxiliary system corrects the monochromatic errors of the mirror. *See* SCHMIDT CAMERA

Advantages. Telescopes, binoculars, and similar instruments are used not only to magnify small objects but also to make them visible if the object contrast is insufficient. The contrast of an image may be defined as

$$K = \frac{I_{max} - I_{min}}{(I_{max} + I_{min})}$$

[illegible] ple, to a [illegible] I_{min} the [illegible] is case, it can be shown that the threshold intensity is proportional to the telescope magnification, provided the pupil of the eye is filled with light.

In viewing objects of low contrast (night vision), a telescope helps the observer by masking out some of the background and background illumination. The effect in this case is, of course, independent of magnification, a tube without lenses would serve the same effect. *See* OPTICAL TRACKING INSTRUMENTS, OPTICS, GEOMETRICAL; RESOLVING POWER (OPTICS). *See also* RADIO TELESCOPE [M H.]

Bibliography: D H Jacobs, *Fundamentals of Optical Engineering*, 1943; A Kuehl, *Z tech Physik*, 1926; R T. Glazebrook, *A Dictionary of Applied Physics*, vol. 4, 1923

Telescope, astronomical

A combination of lenses and mirrors, intended to form real images of extremely distant objects A telescope may be only a single converging lens or mirror but it is more usually made up of several mirrors and lenses The objects studied, such as stars and planets, are so remote that astronomical telescopes are always designed as though the distances were actually infinite, this results in considerable simplification over laboratory instruments dealing with objects at variable distances. A surface, the focal plane, is defined by the position of the images. The focal plane may be a plane surface, but it often is strongly curved, like a cylinder or a sphere. Because telescopes are commonly used as photographic cameras, the shape of the focal plane and the extent of the sky that can be recorded with good images in a single exposure are important considerations.

Simple telescope. In a simple astronomical telescope, an objective lens, mounted at the end of a cylindrical tube, forms real images, A and B, of two stars in its focal plane (Fig. 1). The images of stars may be viewed directly by the eye with some difficulty, but they are more often projected onto a viewing screen, a photographic plate, a photoelectric surface, or other light-sensitive instrument. They may be re-imaged by additional lenses and mirrors as required for a particular type of investigation. For visual observation, a combination of lenses, forming an eyepiece, follows the focal plane. An eyepiece enables the eye to view the star images from a short distance, thus increasing the angle subtended. The stars subtend the angle α_o at the telescope lens, the same angle which they subtend when they are viewed with the unaided eye, but the eyepiece causes the star images to appear separated by the angle α_e when viewed from the image plane. The magnification of the telescope eyepiece combination is the ratio

$$M = \tan \alpha_e / \tan \alpha_o \quad (1)$$

or, equivalently,

$$M = \frac{\text{focal length of objective}}{\text{focal length of eyepiece}} \quad (2)$$

and

$$M = \frac{\text{diameter of objective}}{\text{diameter of image of objective in image plane}} \quad (3)$$

Equation (2) shows that a long-focal-length objective and a short-focal length eyepiece are essential for high magnification. Equation (3) indicates the most direct method for measuring the magnification of a telescope and eyepiece

Resolving power. An upper limit to the useful magnifying power for visual observations is enforced by the wave nature of light. Under certain assumptions, the smallest angular separation (measured in radians) of two stars that can be formed into separate images by a telescope is

$$\alpha = 1.22\ \lambda/D \quad (4)$$

where λ is the wavelength of the light used for the observation, and D is the diameter of the objective. Equation (4) shows that the minimum observable [illegible]

[illegible] standing that high resolving power means a small

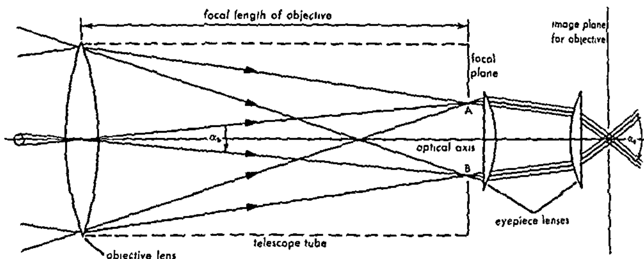


Fig 1. Single-lens refracting telescope

observable angle of separation The smallest angular separation of two objects just viewed as distinct by an average human eye is about 1 minute of arc The angle α in Eq. (4) must be *magnified by a* suitable eyepiece to this value if the eye is to see the smallest detail.

Linear separation. For most nonvisual observation, the linear separation of the star images in the focal plane of the telescope is an important quantity. The linear separation of the images of the two stars A and B is equal to the product of the effective focal length of the objective and the angular distance (in radians) between the stars as viewed in the sky. Because the angular separations of stars can be measured in other ways, the measurement of the linear separation of star images in the focal plane of the telescope leads directly to a determination of the effective focal length of the objective. The linear distance that corresponds to the minimum observable angular separation follows from Eq. (4)

$$\text{Linear resolution} = F\alpha = 1.22\,\lambda\,(F/D) \quad (5)$$

The ratio of the focal length of a telescope objective to its diameter F/D is known as the focal ratio. A large linear scale for a given resolving power can be obtained only with a large focal length.

In addition to its important function in determining the angular positions of the stars, a telescope increases the contrast of astronomical objects with their sky background to a point at which they can be recorded. For example, with a telescope, stars are easily visible in full daylight.

Many different combinations of optical and mechanical elements have been tried in attempts to perfect the accuracy of the angular resolution, which depends mainly on the sharpness of a single star image; and to increase the speed of photographic registration, which depends mainly on the focal ratio.

Seven inherent aberrations that affect the definition (or sharpness), position, and contrast of the image are the five monochromatic aberrations—spherical, comatic, astigmatic, curvature of field, and distortion; and the two chromatic aberrations—longitudinal and lateral (*see* ABERRATION, OPTICAL). These play important roles in the development of the many types of telescopes. It is impossible to build a telescope that is completely free from all aberrations; the different forms of telescopes represent compromises to suit the particular problem at hand.

Refracting telescope. A single-lens telescope is generally satisfactory for astronomical observation if monochromatic light is used, if the ratio of focal length to diameter of the objective is 15 or larger, and if the angular field of view is less than 1°. The coronagraph is a special astronomical telescope that uses a single-element lens (*see* CORONAGRAPH).

Most objectives in refracting or dioptric telescopes are composed of two lenses, a biconvex and a concave, manufactured from glasses so chosen that refraction is obtained with the least possible dispersion (Fig 2). This construction is the simplest one that produces considerable correction of the chromatic aberrations of a single lens, but even the best of two-lens systems show colored fringes around the images. The two-lens objectives can have focal ratios less than 15 and tolerably good angular fields of about 1°.

In general, useful images can be obtained over larger and larger fields, for smaller and smaller focal ratios, as the number of lenses in the telescope objective is increased Objectives with four components are fairly common At focal ratios in the range 5–7, these objectives form usable star images over fields 15° in diameter, but critically sharp definition can be obtained only within the central 5° Still more complicated lens systems

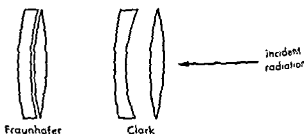


Fig. 2. Achromatic telescope objectives.

been used. The six-element objective with a ratio of 2.5 gives usable, but not first-class, images throughout a 30° field.

eflecting telescope. The reflector, or catoptric, cope can be arranged variously (Fig. 3). Al- gh the catoptric telescope is the most popular because of its perfect achromatism, it exhibits e amounts of other aberrations.

mirror composed of a glass disk, with one con- e spherical surface on which a thin, highly re- ting coating is plated, is satisfactory as a tele- pe for focal ratios greater than 30 and fields of w less than 1°. If the concave reflecting surface a paraboloid of revolution, perfect stellar images e obtained on the optical axis, but they deterio- te rapidly for a wide angular field. The single ncave-mirror telescope that forms its images in e center of the incoming beam of radiation is the pe used for the large, steerable, radio telescopes ith mirror diameters of 30 meters or more.

Because the reflecting telescope returns the eam of incident radiation to a focus within the elescope tube, special design is required to make he image accessible. Small sized receivers that do not obstruct the field of view may be constructed. Tilting the mirror to divert the converging beam to a focal point near one side of the telescope tube may be an alternative if the diameter of the mirror

fall on the main mirror, but it does not change the

reflection at the secondary, the light rays again travel along the central axis of the telescope tube, emerging through the hole in the primary mirror to an accessible focal plane at the rear of the tele- scope. In the Brachyt form, both the primary and

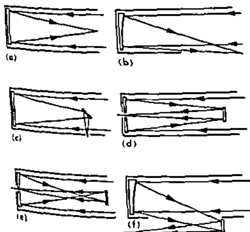


Fig. 3. Usual forms of reflecting telescopes. (a) Prime focus. (b) Herschelian. (c) Newtonian. (d) Cassegrainian. (e) Gregorian. (f) Brachyt.

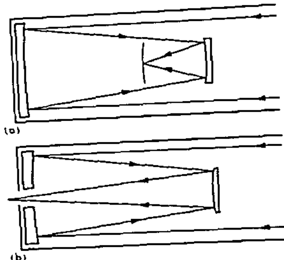


Fig. 4 Catoptric telescopes with extended fields. (a) Schwarzschild. (b) Chrétien.

secondary mirrors are tilted slightly with respect to the light beam. The angles of incidence and the figures of the mirrors can be chosen so that excel- lent definition is obtained over a small field.

Reflecting telescopes have reflecting surfaces that are of simple forms: spheres, spheroids, pa- raboloids, hyperboloids, or planes. The use of these surfaces restricts the angular field of good stellar images. Two-mirror systems using more compli- cated surfaces have been constructed to increase the size of the field of view, two types being shown in Fig. 4 The focal plane of the Schwarzschild

is strongly curved and a plateholder of special de- sign must be used. but this is easily accessible. The Chrétien telescope is short, although focal ra- tios of 6–8 are employed

Refracting-reflecting telescopes. Refractors and reflectors have a number of complementary characteristics, and combinations of the two have resulted in some noteworthy astronomical tele- scopes Generally, the mirror objective is retained because of its perfect achromatism. The only seri- ous outstanding aberration of the paraboloidal pri- mary mirror is coma, an aberration that increases with distance from the optical axis and inversely with the focal ratio. There is a large family of lens systems that will eliminate the imaging errors of paraboloidal, or spherical, mirrors without the in- troduction of other errors in any serious degree. The lens-mirror systems are called catadioptric (Fig. 5).

Reflector with coma correction. The usable field of a paraboloidal mirror can be increased by the addition of a system of coma-correcting lenses near the focus The largest reflecting astronomical tele- scopes are equipped with this kind of corrector Some chromatic and other aberrations are intro- duced by the correctors, but these are negligible compared to the gain in field

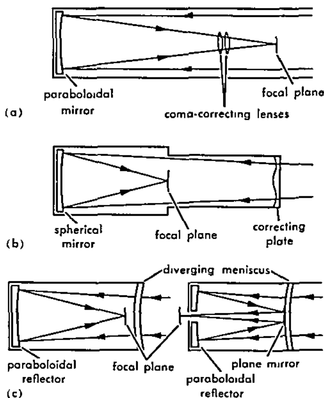


Fig. 5. Reflecting-refracting telescopes. (a) Prime focus reflector with coma-correcting lens. (b) Schmidt telescope. (c) Maksutov telescope.

An interesting catadioptric telescope can be made by forming a reflecting coating on one surface of a properly figured thick lens, thus achieving protection for the fragile reflecting films, but the most satisfactory combinations are those in which the correcting lens is placed in front of the main mirror at about one or two focal distances Two families of instruments of this sort are the Schmidt and Maksutov telescopes.

Schmidt telescope. A Schmidt telescope has a spherical primary mirror that looks through a thin correcting plate placed at twice the focal length of the primary. Near its center, the correcting plate acts as a weak converging lens, near its outer edge, it behaves as a weak diverging lens. The radius of curvature of the focal plane is equal to the focal length of the primary mirror, and a sharply focused field of much greater extent than that of the paraboloidal telescope is obtained Fields of 15° at focal ratios of 1.5 are easily possible with Schmidt systems (*see* SCHMIDT CAMERA).

Maksutov telescope. A similar correction of the principal defects of the paraboloidal reflector can be obtained by replacing the thin Schmidt correcting plate with a weakly diverging meniscus lens. The relatively thick meniscus lens, like the Schmidt correcting plate, is nearly achromatic.

The practice of introducing lens aberrations of just the right type into a reflecting telescope for later correction by the primary mirror is widely applicable and it can be adapted to all of the catoptric telescopes. The catadioptric instruments are designed to provide wide angular fields of nearly perfect images, at low values of the focal ratio, so that photographic exposure times are short.

Comparison of telescopes. In general, refractors are more permanent than reflectors, and offer easier control of stray light. The permanence of the refractor is necessary in programs that require a century or so for their completion (proper-motion studies, positional astronomy). The control of stray light is extremely important in all work where the highest contrast is required (solar investigations, planetary, and lunar studies). However, the maximum size of the refracting telescope is limited by the tolerable absorption in the lenses as they become thicker with increase in diameter, and by the practical difficulty of producing large disks of glass with few flaws. The 1-meter-aperture refractor of the Yerkes Observatory of the University of Chicago is near the upper limit of size.

Reflecting telescopes are inherently achromatic and may be constructed in large sizes. Glass disks can be cast in large diameters; the difficulties of transportation to the telescope site set the practical limit of size. Paraboloidal metallic reflectors for radio telescopes may be assembled in the instrument and can have apertures of some 100 meters or more. Giant telescopes of this kind are important primarily for the collection and concentration of energy from faint objects. Catadioptric telescopes combine the advantage of large size with other valuable characteristics.

Limitations. The primary goal of all astronomical telescope systems is the imaging of celestial objects as precisely as possible. A perfect optical system must be mounted so that the desired part of the sky can be observed and, in most cases, followed automatically. Although such provision has been made, it will be found that the ultimate limit to the performance of telescopes is established by the turbulence of Earth's atmosphere. If this were not so, the 5-meter Mount Palomar telescope would reveal the disks of several stars other than the Sun

The elimination of the effects of atmospheric turbulence is necessary for the perfection of telescopic observations, and two methods have been proposed. The first is the obvious solution, consisting in removing the telescope to a site outside the atmosphere by balloon, rocket, or satellite The second is a solution in terms of correcting the op-

of the future *See* ASTRONOMICAL INSTRUMENTS; SPECTROHELIOSCOPE [R R M.]

Bibliography· G. Z. Dimitroff and J. G. Baker, *Telescopes and Accessories*, 1945, A. G. Ingalls (ed.), *Amateur Telescope Making*, pts. 1-2, 4th ed, 1941, pt. 3, 1953, H C King, *The History of the Telescope*, 1955.

Telestacea

An order of the subclass Alcyonaria. Telestacea are typified by *Telesto* which forms an erect branching colony by lateral budding from the body

ll of an elongated primary or axial polyp. The lon is bandlike or membranous. The sclerites scattered singly, partly fused, or entirely fused form a rigid tube. *See* ALCYONARIA. [K.A.]

eletypesetter

system for automatically operating a linecasting

ting the complex manual operation of a linecasting machine into two simple operations and making one of them fully automatic under the control of perforated tape.

The perforated tape is produced on a Teletypesetter perforator, which has a keyboard layout similar to a typewriter, plus additional keys to control the various functions of the linecasting machine. A tape-punch mechanism perforates combinations containing one to six holes (plus a tape feed hole) in the tape each time a key is depressed. The perforator has a top speed of 900 key strokes per minute. The average operator can easily produce 400 or more lines per hour of 5½- to 9 point type for a 12 pica column width. *See* COMPOSITION (TYPE).

A counting mechanism in the tape perforator automatically counts each character in proportion to its width, and the accumulative total is shown on an indicator scale. This scale clearly shows the condition of the line being composed and tells the operator when the line will justify. Corrections in the tape are made by backspacing and using the rubout key to delete a letter, word, or line as required.

The perforated tape controls the Teletypesetter operating unit, which is attached to the linecasting machine keyboard Thus the linecasting machine is fully automatic under the control of perforated tape produced on the perforator. Production of type is increased by 100–200% over that obtainable by direct normal keyboard operation

As the tape runs through the operating unit, a mechanism "senses" the code perforations in the tape and translates them into lever movements, which cause matrices and spacebands to drop and the assembling elevator to rise when a line is com-

...anical failure occurs.

Teletypesetter equipment is used by most daily and weekly newspapers and by many commercial shops throughout the world By supplementing Teletypesetting equipment with Teletype sending and receiving equipment operating over special long-distance telegraph circuits, the perforated tape can be used to produce type automatically in hundreds of printing plants throughout the country. At present, the two major press associations employ Teletype and Teletypesetter equipment to transmit tape in justified-line form to most of the daily newspapers in the United States. *See* TELETYPEWRITER. [Y.N.S.]

Teletypewriter

An electromechanical device used for transmitting and receiving messages over a telegraph circuit. It is also called a teleprinter. A sending and receiving teletypewriter performs two functions: the keyboard transmitter generates coded electrical signals for transmission over a telegraph circuit; and the typing unit converts such signals into a printed message.

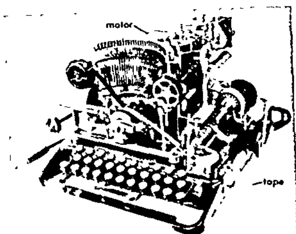


Fig 1 Tape teletypewriter.

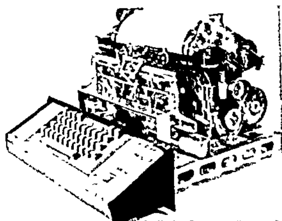


Fig 2. Page teletypewriter

A tape teletypewriter (see Fig 1) reproduces a printed message on a long, narrow tape, approximately 5/16 in wide. This tape is usually gummed on the back so that it can be torn into short strips and affixed to a telegraph blank A page teletypewriter (see Fig 2) prints a message in page form, usually on a continuous roll of paper 8½ in wide.

Baudot code. The Baudot code used in printing telegraphy consists of five code pulses, any one of

which may be either marking or spacing. In single-current signals used for operating most teletypewriters, a marking pulse is an interval of time during which current flows through the circuit and a spacing pulse is an interval during which no current flows. In polar signals, frequently used over long lines, a marking pulse is an interval during which negative current flows and a spacing pulse is an interval during which positive current flows.

When the Baudot code is used to operate teletypewriters, the five code pulses are preceded by a start pulse, which is always a spacing pulse, and

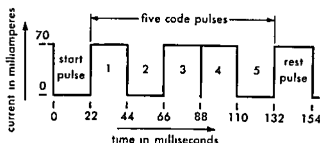


Fig. 3. Five-unit start-stop telegraph code, 7-unit code pattern. This code pulse combination is for the letter F

CHARACTERS				CODE SIGNALS				
PAGE		TAPE						
UC	LC	UC	LC	1	2	3	4	5
		–	A	■	■			
		?	B	■			■	■
		.	C		■	■	■	
		$	D	■			■	
		3	E	■				
			F	■		■	■	
		&	G		■		■	■
#		£	H			■		■
		8	I		■	■		
		BELL	J	■	■		■	
		(	K	■	■	■	■	
		)	L		■			■
.		%	M			■	■	■
,		'	N			■	■	
		9	O				■	■
		0	P		■	■		■
		1	Q	■	■	■		■
		4	R		■		■	
'		¶	S	■		■		
		5	T					■
		7	U	■	■	■		
		;	V		■	■	■	■
		2	W	■	■			■
		/	X	■		■	■	■
		6	Y	■		■		■
		"	Z	■				■
		SPACE				■		
CAR RET.		,	.				■	
LINE FEED		#	=		■			
		FIGURES		■	■		■	■
		LETTERS		■	■	■	■	■
		BLANK						

Fig. 4. Tape and page teletypewriter code assignments, showing only the differences between tape and page codes in the page upper-case column.

are followed by a rest, or stop, pulse, which is always a marking pulse.

There are 32 possible combinations of the five code pulses. One combination is assigned to each of the 26 letters of the alphabet. The blank combination, in which all five code pulses are spacing, is not normally used. The remaining five combinations are used for the following functions. (1) Figures shift, which causes the typing unit to shift to a position to print digits or punctuation marks. (2) Letters shift, which shifts the typing unit to a position to print letters. (3) Carriage return, which causes the printing carriage of a page machine to return from the right margin to the left margin at the end of a printed line. (4) Line feed, used on page machines, which feeds the paper up one line after a line has been printed. (5) Space, which causes the typing unit to space between words. In tape teletypewriters, carriage return and line feed functions are not required, and these two combinations are used for punctuation marks.

Since only 26 code combinations are available for printing in each of the two shift positions, both capital and lower-case letters cannot be used. The upper-case position is reserved for the 10 digits, punctuation marks, and commonly used symbols. One upper-case combination is used for operating an audible (bell) signal.

Operation. An electric motor provides power for operating the keyboard and typing unit. The motor is geared to driving members of clutches on the keyboard shaft and on the receiving shaft of the typing unit. When a key lever is depressed, five code bars are positioned to the right or left (marking or spacing position) in a pattern corresponding to the code combination for the character or function represented by the depressed key lever. A universal bar, operated when any key lever is depressed, actuates a clutch-release mechanism which causes the keyboard clutch to engage and allow a cam sleeve assembly on the keyboard transmitting shaft to start rotating. During this rotation, six cams operate contacts in sequence to generate the start-stop signals. The contacts close to generate a marking pulse and are prevented from closing to generate a spacing pulse. One of these transmitting cams generates the start and rest pulses. The other five cams generate the code pulses corresponding to the selected character. At the end of the revolution of the cam sleeve, the transmitting clutch is disengaged and the cam sleeve remains at rest until a key lever is again depressed.

The start-stop telegraph signals are received by an electromagnet on the typing unit. When current flows through the magnet coils (marking pulse), the armature is attracted to the pole piece of the magnet. When no current flows through the coil (spacing pulse), a spring pulls the armature away from the pole piece. When a start pulse is received, the armature is pulled away from the pole piece. This releases the receiving selector clutch, which then drives a cam sleeve assembly on the receiving shaft. As this assembly rotates, five selector cams

entially position mechanical members on the ng unit either to the marking or spacing posi- , depending on the position of the armature at time each selection is made. After the five code ctions have been set up, a sixth pulse on the sleeve trips another clutch which drives the hanism for printing or performing the selected ction. The character printed or the function formed is determined by the code combination up in the selector mechanism.

Shortly after the fifth code selection is set up, the ecting cam sleeve assembly returns to its rest ition and the receiving selector clutch is dis- gaged. The cam assembly stops rotating and re- ins at rest until the next start pulse is received. e receiving shaft rotates faster than the trans- tting shaft, the most common speed ratios being 7 and 13.12. The receiving shaft makes one rev- ution and returns to its home position while the st pulse is being received. Therefore, the re- ving shaft always comes to rest briefly at the d of a revolution. This insures that the receiving letypewriter always begins each operation in nchronism with the transmitting unit. This start- op synchronization prevents accumulation of any inor speed differences between the sending and ceiving machines.

Only a small portion of each pulse length is re- uired to set up a selection. The remaining length f each pulse provides an operating margin to in- ure correct printing even when the received sig- als are distorted in transmission. A device called range scale allows the instants of selection to be riented with respect to the received signals so that he selections will occur in the middle of each code pulse, even when the signals are badly distorted.

See TELEGRAPHY; TELETYPEWRITER EXCHANGE (TWX) SERVICE; TELEX [F.W.S.]

Bibliography: A. S. Benjamin and W. J. Zenner, A step forward in printing telegraph, Trans AIEE, 73(1):10-15, 1954; C. E. Schultheiss, High-speed teletypewriter equipment for the armed services, Trans AIEE, 73(1):88-93, 1954; F. W. Smith, Modern high speed teleprinters, Western Union ... graph codes, Western Union Tech. Rev., vol. 11 no. 4, 1957.

Teletypewriter exchange (TWX) service

A direct, two-way typewritten communication link between locations throughout continental United States, except Alaska.

TWX uses a terminal machine comparable in ap- pearance and operation to a typewriter and essen- tially like other teletypewriters. Operation is at 60 words per minute (wpm) using the conventional start-stop teletypewriter code consisting of a 1-unit start pulse, 5 units (bits) for information and a 1.4-unit stop pulse, a total of 7.4 units for each character. Messages typed on one machine are si- multane... ...ed on any other teletype- writer to which it is connected through the ex- change service facilities. See TELETYPEWRITER.

Operating procedure. TWX has more than 45,- 000 subscribers, any one of whom can call any other. To do so, a subscriber turns on his machine, which automatically signals the operator. The function of the operator is comparable to that of a manual telephone operator, except that she uses her teletypewriter, rather than voice, to communi- cate with the calling and called stations and com- plete the connection. Since mechanization began in 1938 more than half of the approximately 275 switching centers have been converted to automatic switching.

After connections are established by the opera- tor, the subscribers use their keyboards to commu- nicate directly with each other on a chit-chat basis. Alternatively, at stations where the machine in- cludes a tape unit, transmission may be from punched paper tape prepared in advance. Because transmissions are received in exactly the form in which they are sent, tabulations and page format control are practicable. Also available is confer- ence service, providing simultaneous connection of a number of stations, and unattended service, which permits receipt of communications when the receiving teletypewriter is unattended, as for example, outside of office hours.

Some of the most common uses of TWX are to transmit orders, specifications, and shipping in- structions; to interchange information as to market conditions, availability, and price of goods; to make purchases and sales; to interchange credit in- formation; and to interchange information in con- nection with crimes, arrests, and police orders.

Comparison with other services. Switching ar- rangements provided for TWX are comparable, sometimes identical, to those provided for tele- phone. TWX employs the same fundamental prin- ciples of transmission as other teletypewriter serv- ices, the circuits and terminal equipment units are also comparable. However, the service is fundamen- tally different from other classes of telegraph serv- ice.

Compared with private-line teletypewriter serv- ice, TWX connections are established when desired in periods as short as 3 minutes; private lines are usually established either on a permanent basis or for previously contracted periods. Interconnection is available with any subscriber to the service; with private lines, generally, only the stations of the same customer may be interconnected.

Compared with message telegram service, TWX provides direct communication between the prem- ises of customers. Only the customers (or their employees) take part in the handling of the com- munication. Telegram service is a pick-up and de- livery service involving the relaying of a message from sender to addressee by the telegraph com- pany. TWX is a two-way service permitting inquiry and reply during the same connection; telegraph service is one-way. TWX is available only between subscribers to the service and requires the instal-

lation of a TWX teletypewriter on the subscriber's premises. Telegram service is available to the general public for communication with anyone.

In many other countries, principally Europe, TELEX service is provided. I[illegible] same as [illegible] are that [illegible] (for establishing connections), machine speeds are 66 wpm, and about 12 upper-case keyboard characters are not the same (such as £ and $). All of the numerals, alphabet, and most commonly used punctuation marks are the same. TELEX subscribers may communicate with TWX subscribers through the international common carriers. *See* TELEGRAPHY; TELEX. [V.N.V.]

Television

The electrical transmission and reception of transient visual images. Like motion pictures, television consists of a series of successive images, which are registered on the brain as a continuous picture because of the persistence of vision. Each visual image impressed on the eye persists for a fraction of a second. In television in the United States, 30 complete pictures are transmitted each second, which with the use of interlaced scanning is fast enough to avoid evident flicker.

At the television transmitter, minute portions of a scene are sampled individually for brightness (and color for color television) and the information for each portion is transmitted consecutively. At the receiver, each portion is reproduced in its proper position and with correct brightness (and color) to reproduce the original scene.

The scene at the transmitter is focused on a photoelectric screen of a camera tube. Each portion of the screen is changed by the photoelectrons to a degree depending upon the brightness of the particular portion. The screen is scanned by an electron beam just as a reader scans a page of printed type, character by character, line by line. When so scanned, an electric current flows with an instantaneous magnitude proportional to the brightness of the portion scanned. *See* TELEVISION CAMERA TUBE.

Variations in the current are transmitted to the receiver, where the process is reversed. An electron beam in the picture tube is varied in intensity (modulated) by the incoming signals as it scans the picture-tube screen in synchronism with the scanning at the transmitter. The photoelectric surface of the picture tube produces light in proportion to the intensity of the electron beam which strikes it. In this way the minute portions of the original scene are recreated in their proper positions, brightness, and (for color transmission) color values. The elements of a television system are shown in Fig. 1. For more detailed discussion, see TELEVISION RECEIVER, TELEVISION TRANSMITTER.

In the United States an individual picture (frame) is considered to be made up of 525 lines, each line containing several hundred picture elements. All these lines are scanned and the light values are sent to the receiver in 1/30 sec so that each second 30 pictures are received. T[illegible]

[illegible] the scanning beam is directed to the next

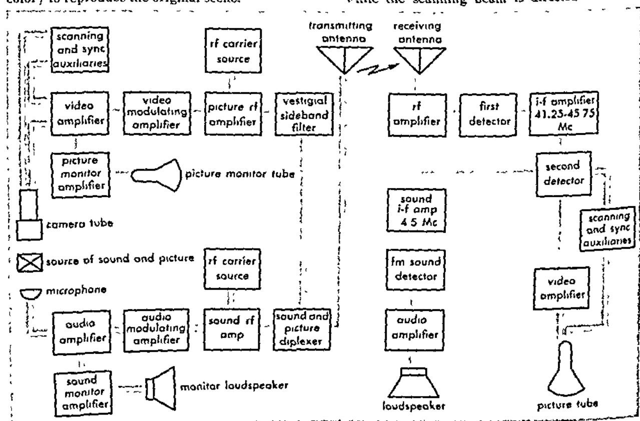


Fig. 1. Elements of television system. (*From K. Henney, ed., Radio Engineering Handbook, 5th ed., McGraw-Hill, 1959*)

line. During these short intervals, synchronizing signals are transmitted to keep the scanning process at the receiver in step with that at the transmitter.

Scanning. To take full advantage of the persistence of vision, each frame is scanned twice with alternate lines being scanned in turn. This technique is called interlaced scanning.

Since 525 horizontal lines are scanned in 1/30 sec, the horizontal scanning rate for black and white pictures is 15,750 times per second. Since two vertical fields are scanned in 1/30 sec, the vertical scanning rate is 60 times per second See TELEVISION SCANNING.

Bandwidth. The bandwidth required for any information transmission system is a function of the number of bits of information, or the detail, to be transmitted per second. For a television picture, the greatest detail would be required if the picture consisted of a checkerboard pattern of the smallest squares the system must handle to provide acceptable resolution The standard of 525 lines sets the vertical detail and the standard aspect ratio (picture width to height) of 4/3 requires 700 horizontal picture elements for equal horizontal and vertical resolution, or 350 sets of alternate black and white squares. The picture is reproduced 30 times a second, for a total of 525 × 350 × 30 or 5,512,500 complete cycles per second

Less detail than this is actually transmitted and received, and the highest video frequency actually transmitted is 4.2 Mc. See BANDWIDTH REQUIREMENTS (COMMUNICATIONS)

Frequency. The band of frequencies assigned to a television station for the transmission of synchronized picture and sound signals is called a television channel. In the United States a television channel is 6 Mc wide, with the visual carrier frequency 1.25 Mc above the lower edge of the band and the aural carrier 0.25 Mc below the upper edge of the band.

Television channels in the United States are identified by numbers, starting with channel 2. The frequency assigned to channel 1 was later reassigned to other uses The table shows that these channels are in three frequency bands Channels 2-6 occupy the region from 54 to 88 Mc, channels 7-13 are from 174 to 216 Mc, and channels 14-83 are from 470 to 890 Mc. The first two groups of channels fall in the very high frequency (vhf) band, the channels in the last group are in the ultrahigh frequency (uhf) band.

Sound transmission. In the United States, the sound portion of the program is transmitted by frequency modulation at a carrier frequency 4.5 Mc above the picture carrier. Maximum frequency deviation (bandwidth) of the sound signals is 25 kilocycles (kc).

The normal frequency response is altered in the transmitter to emphasize the higher audio-frequencies with respect to the lower frequencies. Called preemphasis, this is accomplished by a circuit that causes the audio response to increase with frequency A corresponding circuit is used in the

Television channels in the United States

Channel no	Frequency band, Mc	Channel no	Frequency band, Mc
2	54-60	43	644-650
3	60-66	44	650-656
4	66-72	45	656-662
5	76-82	46	662-668
6	82-88	47	668-674
7	174-180	48	674-680
8	180-186	49	680-686
9	186-192	50	686-692
10	192-198	51	692-698
11	198-204	52	698-704
12	204-210	53	704-710
13	210-216	54	710-716
14	470-476	55	716-722
15	476-482	56	722-728
16	482-488	57	728-734
17	488-494	58	734-740
18	494-500	59	740-746
19	500-506	60	746-752
20	506-512	61	752-758
21	512-518	62	758-764
22	518-524	63	764-770
23	524-530	64	770-776
24	530-536	65	776-782
25	536-542	66	782-788
26	542-548	67	788-794
27	548-554	68	794-800
28	554-560	69	800-806
29	560-566	70	806-812
30	566-572	71	812-818
31	572-578	72	818-824
32	578-584	73	824-830
33	584-590	74	830-836
34	590-596	75	836-842
35	596-602	76	842-848
36	602-608	77	848-854
37	608-614	78	854-860
38	614-620	79	860-866
39	620-626	80	866-872
40	626-632	81	872-878
41	632-638	82	878-884
42	638-644	83	884-890

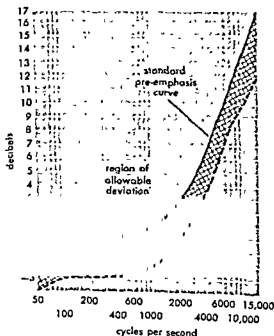


Fig 2. Standard preemphasis curve

receiver to produce an equal and opposite decrease of response to higher audio-frequencies. By so doing, noise produced in the receiver is attenuated without the over-all system audio-frequency response being affected. The Federal Communications Commission (FCC) requires that the response of the aural transmitting system must not exceed the limits shown in Fig. 2.

The harmonic distortion of the audio-frequency signals must not exceed the following rms values when the harmonics are measured out to 30 kc:

Frequency range, cps	*% Distortion*
50–100	3.5
100–7500	2.5
7500–15,000	3.0

Picture transmission. The visual signals are transmitted at a carrier frequency 1.25 Mc above the lower limit of the channel, using amplitude modulation and vestigial sideband transmission (*see* AMPLITUDE MODULATION). The upper sideband is fully transmitted, but the lower sideband is attenuated beginning at 0.5 Mc below the carrier. Attenuation is virtually complete at 1.25 Mc below the carrier. This method of transmission reduces the required bandwidth of the channel and allows more channels to use the available space in the radio spectrum. Figure 3 is an output characteristic of a transmitter for a channel 2 station, showing how the 6-Mc band is used for picture and sound transmission.

Negative modulation for picture transmission is used in the United States to minimize the effects of noise during synchronizing signal reception. In negative modulation an increase in brightness causes a decrease in transmitted power. Some foreign systems use positive modulation.

Ghost image. Radio waves from a transmitter to a receiver normally follow a straight path. However, it is possible for such waves also to follow a longer path, by reflections from large objects, such as mountains or buildings. The reflected waves arrive later than the direct waves, so a second picture is reproduced from a fraction of an inch to several inches to the right, depending on the length of the indirect path. The second picture is called a ghost image. There may be several such ghosts when reflecting objects are in an area.

Snow. Noise voltages are produced in the input circuits of all radio receivers by thermal agitation in wires and by the uneven flow of vacuum tube currents. When the signals from a television transmitter are too weak to overcome this noise, it produces a display that looks like and is often called snow.

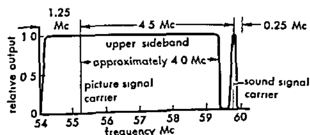


Fig. 3. Output characteristic of a television transmitter. (From K. Henney, ed., *Radio Engineering Handbook, 5th ed.*, McGraw-Hill, 1959)

Scrambling of television signals. If the synchronizing signals are caused to vary at random but predetermined rates, the signal is unusable in conventional receivers. Special receivers may contain unscrambling devices which, by a specially transmitted signal or a built-in keyer, cause the receiver to synchronize and produce normal pictures. Scrambling devices were developed for paid television, sometimes called tollevision, to prevent reception by anyone but a subscriber. *See* CLOSED-CIRCUIT TELEVISION.

Recording television programs. Television programs are recorded for rebroadcast at a later time and for many other reasons. There are two principal methods: kinescope and tape recording.

Kinescope recording. In this method the images produced on a television picture tube are photographed on motion picture film. Also known as video recording and teletranscriptions, this technique is highly developed and is relatively inexpensive. It is used for various purposes, including delayed broadcasting, syndication, record purposes, and future production planning. A processed negative may be used for immediate transmission by reversal of circuit polarity and other changes in a scanning system.

Video tape recording. In this technique television sound and pictures are recorded on magnetic tape. Electrical signals constituting the sound and picture pass through magnetic recording heads. As the tape is drawn across these heads at a speed of 15 in./sec, the signal currents produce magnetization of finely powdered iron in the tape emulsion. The signals are recovered when the magnetized tape is drawn across a reproducing head and the magnetic fields intercept a pickup coil.

Video tape recording (VTR) has the great advantage that it may be used immediately without further processing, and the tape may be erased by demagnetization and reused. It is of unique value for prerecording programs, reruns, special events coverage, test programs, and delayed program transmission for different network station time zones.

The tape consists of a coated plastic base 2 in. wide. A 12½-in. reel provides 64 min of program time. Prior to recording, the video signal frequency modulates a carrier of from 4.25 to 5.9 Mc. A vestigial sideband signal is recorded on transverse tracks 0.010 in. wide by means of four magnetic heads equally displaced on the circumference of a 2-in. wheel rotating 14,400 times per second. About 18.4 scanning lines are recorded on each transverse track. One picture frame (525 lines) comprises 32 transverse tracks or ½ in. of tape length. The accompanying sound and control signals are recorded on separate longitudinal tracks by means of separate recording heads. In reproduction the

ecovered from the tape and demodu-

elevision service. The FCC has es- definition two grades of television e United States Grade A service pro- ely high freedom from interference stations and also good free-

the best 70% of all receiver locations r geographical limits of this service. rvice recognizes that service is provided e more vulnerable to interference and pecifies that equal service is available, y 50% of all receiver locations at the istance. For other aspects of television TELEVISION; TELEVISION NETWORK, N STANDARDS; TELEVISION STUDIO.

[R.F.GU.]

4 453–458, 1957; D. G Fink (ed.), ... *Engineering Handbook*, 1957, C. P Gins-

ision camera

elevision camera and its auxiliary apparatus, red to translate the optical image of the scene televised into a video signal suitable for mission, form what is called a camera chain e camera is kept as small as possible for flexi- y and ease of handling. Only necessary elec-

plies, and monitoring facilities, are usually in- lled in the control room. A synchronizing gen- ator provides the "sync" pulses used in the mera and camera control unit to create a video gnal.

Monochrome cameras. The typical monochrome amera contains the lens system and optical focus- ng control, the image orthicon camera-tube as- embly (including deflection coils, focusing coil and alignment coils), a high-voltage power supply for the image orthicon, the deflection chassis, the blanking amplifier, the video preamplifier, and blowers and heaters which maintain the correct operating temperature for the camera tube. Most cameras are also equipped with an electronic view-

High quality, color corrected lenses are used to focus an image on the photosensitive surface of the image orthicon tube. Optical focusing is achieved by moving the lens with respect to the face of the camera tube or by moving the camera tube with respect to the lens

The deflection chassis contains the circuitry for the generation of the current waveforms which, when applied to the deflection coils, produce the linear scanning motion of the electron beam in the image orthicon tube. The deflection circuits are synchronized by pulses supplied by the sync generator, called horizontal and vertical driving pulses. Under the influence of the scanning fields, the electron beam in the image orthicon scans out a prescribed pattern (called a raster) on the glass target

see TELEVISION CAMERA

vertical driving pulses to form a blanking ... which causes the target of the image orthicon tube to assume a sufficiently negative potential to cut off the signal output of the image orthicon. The blank- ... during the time the electron

from the bottom of the raster to the top of each vertical field. Since no signal is transmitted during the retrace periods, the retrace lines are not visible to the home viewer.

The image orthicon tube requires several electrostatic fields for its operation. The required potentials are supplied by a pulse-type power supply, which is driven by pulses appearing across the horizontal deflection coils The image orthicon focus coil establishes a magnetic field along the axis of the tube, which, in conjunction with the electrostatic fields, establishes the conditions necessary to control the electron paths within the tube. The current through the focus coil is electronically controlled to remain substantially constant The alignment coils establish electromagnetic fields, which

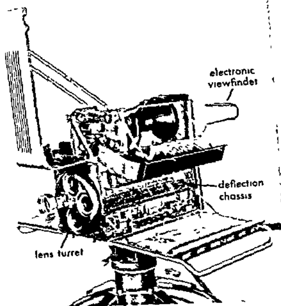


Fig 1 Monochrome camera. (*Radio Corporation of America*)

correct any slight deviations in the initial direction of the electron beam.

The video preamplifier raises the signal output of the image orthicon from approximately 0.05 volt to about 0.5 volt and also provides a 1.0-volt signal for driving the viewfinder. A typical video preamplifier has about eight stages of amplification. The amplifier has compensation circuits which keep the amplification essentially constant over a band of about 8 megacycles (Mc). The preamplifier is required because the output signal from the image orthicon is too low to be transmitted over long cables to the control room without crosstalk or interference.

Figure 1 is a view of an opened monochrome studio camera. The lens turret is controlled by a handle at the cameraman's position. The controls located at the camera are used to set the operating potentials of the image orthicon tube, and to adjust scanning height, width, centering, linearity, and the output level of the camera signal. These and other auxiliary controls are normally adjusted only during camera setup.

Camera control unit. The essential circuits in the control unit include a video amplifier with compensation circuits to amplify and process the video signal, a monitoring kinescope to provide a picture of the camera output for checking picture quality and focus, and a cathode-ray oscilloscope to measure the signal level and observe the waveform of the component parts of the video signal.

Color cameras. One of the basic requirements for color television is the separation into three component parts of the optical image picked up by the camera. This is done by special color filters, which provide separate red, green, and blue images These three images are directed to three image orthicon tubes. Each tube has the same basic associated circuitry as the single-tube monochrome camera, plus additional circuits in the camera and camera control unit to meet the requirements of the color system.

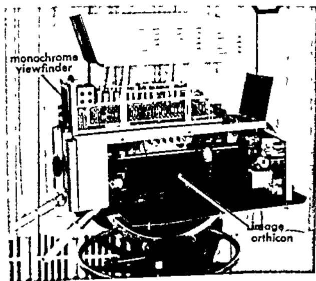


Fig. 2. Typical color camera. (*Radio Corporation of America*)

Figure 2 shows an opened color camera. The camera is usually mounted on a pedestal for easy handling. The viewfinder on top of the camera provides a monochrome picture for the cameraman's use.

Figure 3 is a sketch of the optical system of a color camera. The light splitting is done by dichroic mirrors, which transmit light energy in one portion of the spectrum and reflect light energy in other portions of the spectrum. The condenser lens and the relay lens serve to lengthen the optical path to provide the space necessary for the light splitter. Astigmatism is introduced by the dichroic mirrors because the displacement of the rays passing through the mirrors differs in the horizontal and vertical directions. Correction for astigmatism is provided by use of additional flat glass plates.

In a camera system the outputs of the three color channels are electrically combined to produce a single complex signal. Careful registration of the three optical images in the camera is required to reproduce accurately the color and detail at each point in the scene. Rather stringent design requirements are imposed on the deflection system of the camera, since all three rasters scanned on the targets of the three separate image orthicon tubes must be identi-

[illegible]

optical system must likewise be [illegible] adjustment.

The color camera control units have the circuits and control for the processing of the three color signals. The basic circuits are similar to those used in the monochrome cameras except that nearly all controls are in triplicate

Colorplexer (encoder). The signals produced by the three color channels are luminance signals corresponding to the brightness of each of the primary color components of the scene. To form a compatible color signal capable of being received by both color and monochrome receivers, these three signals are combined in precise proportions in a matrix—a combination of resistors and amplifiers which effectively adds and subtracts the required proportions of the three camera signals to produce a new set of three signals. One component represents the addition of the three camera signals in the correct proportions to provide a luminance signal of the over-all scene This signal is essentially the same as that provided by the monochrome television system The other two signals, termed I and Q signals, are fed to the multiplexer section of the colorplexer and are used to modulate the 3.58-Mc color subcarrier in a two-phase modulation system. The output of the two modulator stages contains color information which is added to the luminance signal and color synchronizing signals to form the color video signal *See* COLOR TELEVISION

Film reproduction. The projectors used for television film reproduction use what is called a 3 2 intermittent mechanism to adapt the 24-frame motion picture frame rate to the 30-frame television system. In this mechanism one frame of the motion picture film is held in the projector gate for two

Fig 3 Optical system of a color camera. (Radio Corporation of America)

elevision fields or 2/60 sec, the next frame is held for three fields or 3/60 sec, and this sequence is repeated continuously. Two film frames require 1/12 sec, which is equal to the time of five television fields. This preserves the 24 frames/sec film rate

[illegible]

it occurs once in each television field during the interval when the film is stationary in the gate. A mirror system is sometimes used to permit one film

[illegible]

with its deflection coils, focus and alignment coils, a video amplifier; and a blanking amplifier The optical image provided by the projector is focused on the photosensitive surface of the vidicon tube. The light intensity of the image reaching the vidicon tube may be adjusted by an adjustable neutral density filter, which changes the light intensity without materially altering the spectral characteristics of the light.

The camera control unit has a video amplifier with compensation circuits to give the best film reproduction A polarity-changing circuit is included so that a positive picture can be obtained from either positive or negative film, as required The controls for the vidicon circuits and for adjusting the video signal to meet the transmission standards are part of the camera control equipment. The power supplies and the deflection amplifiers are also located in the film control room

Color film reproduction The optical system of the color film camera uses a color splitter essentially the same as that used in the color studio camera. By passing light through color filters, red, green, and blue images are provided, one for each of the three vidicon tubes.

The circuits in the color film camera are substantially the same as in the monochrome camera, except that the registration requirement demands circuits for deflection which are inherently stable and have excellent linearity

The color film camera control has circuits for controlling the three vidicon tubes, and processing amplifiers, which are required to establish a high-quality color picture. A color monitor and a cathode-ray oscilloscope are used for checking the picture quality, establishing the required video levels, and examining the video wave forms. The color film system uses a colorplexer for combining the three camera signals to produce the compatible color video signal

Field equipment. Programs originating away from the studios are usually covered by the use of mobile units. Field equipment for monochrome use is designed so that it can be separated into several units for easy handling For example, the camera and viewfinder are separate units. The camera control units, the switching unit, the television monitors, and audio equipment usually are installed in the mobile unit in a semipermanent manner. The mobile unit serves as a control room for many types of pickups Color mobile units usually use studio equipment which, with the exception of the cameras and accessories, is permanently installed. The video signal from the mobile unit is transmitted to the studio or transmitter by microwave relay or by the use of wideband cables. [C K C.]

Bibliography · Color Television Issue, *Broadcast News* [illegible]
ples ar [illegible]
Color [illegible]
ski, 3 Vidicon color film camera, *Broadcast News* No 77, May–June 1954; J. W. Wentworth, *Color Television Engineering*, 1955.

Television camera tube

An electron tube that converts an optical image into an electrical television signal. The tube is used in a television camera to generate a train of elec-

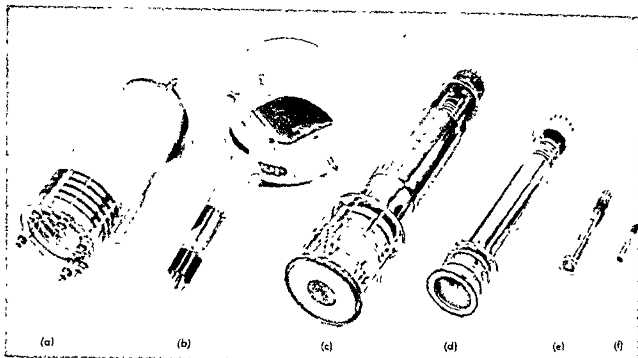


Fig. 1. Some typical camera tubes. (a) Image dissector, (b) iconoscope; (c) 4½-in. image orthicon; (d) 3-in image orthicon; (e) 1-in. vidicon; (f) miniature vidicon (developmental).

trical pulses representing the light intensities present in an optical image focused on the tube. Each point of this image is interrogated in its proper turn by this tube, and an electrical impulse corresponding to the amount of light at that point of the optical image is generated

Television camera tubes are designed primarily to pick up live programs, indoors or outdoors, as well as to reproduce motion pictures and other filmed material. Tubes developed for these purposes are the iconoscope, the image iconoscope, the orthicon, cathode-potential-stabilized (c-p-s) emitron, and the image orthicon. Less complex or smaller camera tubes have been developed for industrial or closed-circuit television service These are the vidicon and the image dissector The former also is finding wide use in broadcast television Figure 1 shows a group of typical television camera tubes.

Although the television camera tube is sensitive primarily to visible light, special tubes have been designed that are sensitive to radiant energy in the infrared, the ultraviolet, and the x-ray portions of the electromagnetic spectrum. The x-ray type has found use in industrial nondestructive testing systems, where the tube replaces a film and produces an immediately available x-ray picture. In medicine, this type of tube can be used instead of a fluoroscopic screen to transmit a picture through a closed circuit to a remote viewer. This arrangement makes it unnecessary for the examiner to adapt his eyes to the dark and reduces both his and the patient's exposure to x-rays.

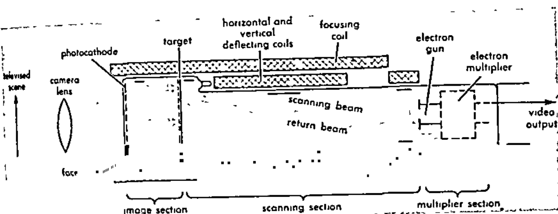
Vidicons sensitive to ultraviolet are being made for high-resolution microscopes. They are also utilized to view the structure and metabolism of living cells, which are for the most part transparent to visible light. For use in military surveillance work, image orthicons have been devised that operate in starlight, and vidicons have been developed that are sensitive to infrared radiation.

The major classification of camera tubes is based

that point is being interrogated. In a storage tube an electric charge accumulates at each point during the interval between successive scans. Because the storage-type tube uses the electric charges generated by the light during the comparatively long intervals between successive scans of the image, it is more efficient and more sensitive. Such other types of television signal generators as flying-spot and monoscope devices are not classified as television camera tubes because the electron tubes used do not convert an optical image into an electrical television signal

Storage-type cameras are further classified according to whether the light-sensitive element utilized is photoemissive or photoconductive. When photoemissive materials absorb light they emit electrons. When photoconductive materials absorb light their electrical conductivity changes.

Image orthicon. Perhaps the most complicated camera tube, the image orthicon has exceptionally high sensitivity and the ability to handle a wide range of light values and contrasts. For these and other reasons, the image orthicon is used almost exclusively in studio and outdoor broadcast television in the United States and many other countries. It is also used in sets of three in color television cameras, each tube generating a signal representing respectively the red, green, and blue components of the light.



[Fig.] 2 The image orthicon and its associated deflect[ing] and focusing coils. (From D. G. Fink, ed., Televi[sion] Engineering Handbook, McGraw-Hill, 1957)

[T]he image orthicon is divided into an image sec[tion], a scanning section, and a multiplier section [(Fig.] 2). These can be considered as three tubes [with]in a single vacuum envelope.

[I]mage section The photoemissive layer is de[posi]ted as a continuous film inside the tube face[pla]te. This layer is called a photocathode and is

[...]e. As this layer is a fairly good electric conduc[tor], a charge pattern does not build up on its sur[fac]e as electrons are emitted When the image or[thi]con is in operation, a light image is focused on [th]e photocathode, whose electrons absorb the en-

[...]arget.

The target consists of two structures (Fig. 3) [f]acing the photocathode and stretched tightly across the end of the image section is a wire mesh screen (over 500,000 openings per square inch) called the target mesh. The mesh is placed several thousandths of an inch from the second structure, the target glass, a glass membrane less than .0002 in thick Most of the photoelectrons pass through the target mesh and hit the target glass. Each photoelectron has several hundred volts energy and knocks several additional electrons from the target glass surface, producing a positive charge at the impact point (*see* SECONDARY EMISSION). The pattern of positive charges corresponds to the light image falling on the orthicon's faceplate. The secondary electrons leaving the target glass are collected by the target mesh, which is held at a slightly more positive voltage.

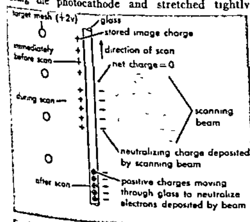


Fig. 3 Image orthicon target structure showing charging and discharging action (From D G Fink, ed, Television Engineering Handbook, McGraw-Hill, 1957)

Scanning section The positive charge pattern is stored on the target glass, which divides the image section from the scanning section. A beam of low-velocity electrons generated by an electron gun is made to scan the rear surface of the glass by varying magnetic fields within the tube. As the beam moves across the glass it deposits electrons wherever positive charges have been built up on the image side The resistance of the glass is carefully controlled so that charges can move from one face to the other before the scanning beam returns to the same spot, yet the glass is designed to insulate well enough to inhibit lateral movement of the charges, which would alter the pattern of the charge image When enough electrons are deposited by the scanning beam to neutralize the charge on the glass and reduce it to the potential of the electron gun cathode, the electrons following thereafter turn about and return to the electron gun When the beam scans an uncharged (dark) area, the full beam is returned. When the beam scans a highly charged (bright) area, most of the beam is deposited and little returns. Thus the beam is amplitude modulated. The variations in the return beam current constitute the television picture information, at low intensity The return beam is amplified about 1000 times in the electron multiplier sec-

tion of the tube and is taken out at the anode of the multiplier as a video signal current.

Multiplier section. The electron multiplier is of unique construction, although it operates like the multiplier used in a multiplier phototube. It consists of a flat first-dynode structure, which is also the screen grid of the electron gun, and a series of pinwheel multipliers. When the return beam strikes the first dynode, a shower of secondary electrons cascades through the pinwheels, where repeated secondary emission multiplies their number. The final group of electrons is collected by the anode and forms the signal current. This current flows across a load resistor in the anode circuit, developing a voltage that is fed to a video amplifier for further amplification. See PHOTOTUBE, MULTIPLIER.

The image orthicon is made in several models for black-and-white television cameras, color television cameras, and cameras designed to operate at low light levels. The difference among these tubes is chiefly in the capacitance of the storage element (the target), which is adjusted by changing the spacing between the target mesh and the target glass.

Iconoscope. The first practical storage-type tube developed, the iconoscope was used in early live television broadcasting but is now used only in motion picture reproduction. It has since been supplanted in the studio by the image orthicon or the c-p-s emitron. The vidicon is replacing the iconoscope as a film reproducer

The iconoscope contains a target and, unlike the image orthicon, an electron gun located on the front side of the target (Fig. 4) The target consists of three elements· a conductive back or signal plate, an insulating support, and a photoemitting surface The latter surface consists of a mosaic of light-sensitive cesium-silver oxide, which is a good electrical conductor The "islands" of this mosaic are distributed at random but are separate and insulated from each other The insulation enables them to store the positive charges that result when light falls on them and electrons are emitted. The scanning beam bombards the mosaic surface with hi velocity electrons, neutralizing the stored posit charges. In addition, the scanning beam electro expel to the higher-voltage collector a consideral number of secondary electrons, more from the are where light does not strike the mosaic surface a less from the areas where the light causes a po tive charge. The secondary electrons continue leave as the scanning beam strikes until the ar under the beam reaches the voltage of the colle tor, the highest-voltage electrode in the tube.

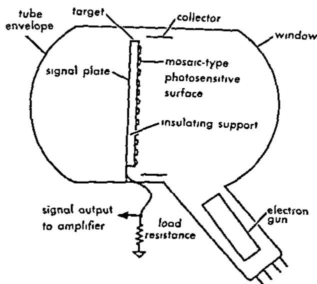


Fig. 4. The iconoscope.

The mosaic is electrically coupled to the sign plate by electrostatic capacitance between the tw Therefore, changes in the electron current flowin away from the mosaic as the beam travels fro heavily charged (lighted) to weakly charged (un lighted) areas cause corresponding changes in th current flowing in and out of the signal plate lead This current, flowing through the load resistance i the signal plate circuit, produces a signal voltag which is amplified to form the television video sig nal.

The picture signal developed by the iconoscope tube is not precisely representative of the image as not all of the secondary and photoemitted electrons go to the collector Some fall as a gentle rain on the mosaic and discharge some of the charged areas, creating random shadows and highlights in the reproduced picture. This defect is characteristic of the iconoscope's high-velocity scanning process, other camera tubes employ a low-velocity scanning beam to minimize this effect.

The iconoscope requires a lot of light to operate properly, is a rather bulky tube to use, and is difficult to operate. However, it was of prime importance in the early development of commercial broadcast television.

Image iconoscope. The image iconoscope is an outgrowth of the iconoscope tube. It has a moderate sensitivity and has been refined to minimize some of the troubles that plague the iconoscope because of its high-velocity electron scanning beam It is still used in some studio television cameras in Europe.

In this tube the photosensitive layer is not part of the image storage element but is a continuous semitransparent layer deposited on the faceplate of the tube, similar to the photocathode of the image orthicon (Fig 5). The target, which is the storage layer, is parallel to the faceplate and a few inches away from the photosensitive layer. These two elements comprise the image section of the tube. When a light image is focused on the photosurface, electrons are emitted. A high-voltage field produced by electrodes within the tube draws the electrons away in streams resembling the bristles of a brush. The image-focusing coils act as an electron lens, focusing these streams to sharp images on the target

The target is an insulator with its surface treated to make it a secondary emitter. Where the photoelectrons strike, a number of secondary electrons are emitted and a positive charge is produced. The image section can be [illegible] d as a device

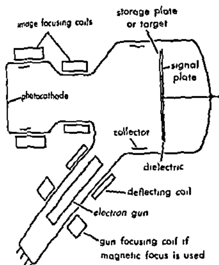


Fig. 5 The image iconoscope tube and its associated deflecting and focusing coils. (From D G Fink, ed., *Television Engineering Handbook*, McGraw-Hill, 1957)

for transferring an electron pattern from a photosensitive surface to an insulating layer for storage The stored positive charges duplicate the light pattern of the optical image.

The signal output is generated by a high-velocity electron beam as it *scans the stored charge pattern* on the target surface. The method of signal generation *from this point* on is identical to that of the iconoscope.

The *image* iconoscope incorporates a number of features to prevent the secondary electrons, produced by the scanning beam and by the photoelectron streams, from dropping back onto the target As a result, the picture is fairly precise and relatively free of unwanted distortions. *The imaging*

studio broadcast service. It is not used where the lighting level and light contrast cannot be closely *controlled, such as in outdoor camera* work or industrial television situations.

The important portion of this tube is the target, which consists of a semitransparent photoemitting mosaic, an insulating support for the mosaic, and a transparent signal plate on the opposite side of the insulator. The mosaic elements and the signal plate are *capacitively coupled by the dielectric support* (Fig 6).

When a scene is imaged on the mosaic, photoelectrons are emitted These electrons are attracted toward the positive electrodes at the electron gun. A pattern of positive electric charges is left on the myriad mosaic elements The scanning beam detects *the positive-charge pattern by depositing low-velocity* electrons on each element until it reaches the zero potential of the cathode of the electron gun. As electrons are deposited on these charged areas, an equal number of electrons flows out of the capacitively coupled signal plate.

A stabilizing mesh in the tube is maintained at *15–20 volts to prevent any element of the mosaic* from charging up beyond this voltage This procedure prevents the scanning beam electrons from landing with enough voltage to eject more secondary electrons than the number of electrons that strike the element. If the latter occurred, the mosaic would charge up uncontrollably and ruin the *picture*

Cathode-potential-stabilized operation refers to the use of a stabilizing mesh and *low-velocity* electrons in the scanning beam to maintain the potential of the mosaic at cathode potential

Vidicon. The vidicon is a small television camera tube that was developed primarily as a closed-circuit or *industrial television camera tube.* Although the name initially applied to a particular 1-in.-*diameter camera tube 6½ in. long,* it has been generally applied to a number of camera tubes employing a photoconductive light-sensitive surface.

Nearly all closed circuit television cameras utilize a vidicon Its small size and the simplicity of operation and adjustment make it well suited for use in systems to be operated by relatively unskilled people Because of the precise and *sharp picture* that it can develop, it is also used in television

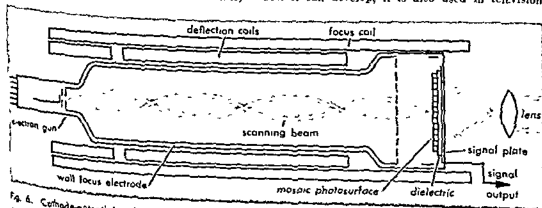


Fig. 6. Cathode-potential stabilized *emitron and the* associated deflecting and focusing coils. (From D G Fink, ed., *Television Engineering Handbook*, McGraw-Hill, 1957)

broadcast service, primarily for the reproduction of motion picture films. The vidicon tubes have a moderate sensitivity, comparable to that of commonly used motion picture film, and can be used in most industrial locations without auxiliary lighting. The speed of response, or ability to capture motion, is at present somewhat less than that of some other television camera tubes.

The vidicon is a simply constructed storage type of camera tube (Fig. 7). The signal output is developed directly from the target of the tube and is generated by a low-velocity scanning beam from an electron gun.

The target consists of a transparent signal electrode deposited on the faceplate of the tube and a thin layer of photoconductive material which is deposited over the electrode The photoconductive layer serves two purposes. It is the light-sensitive element, and it forms the storage surface for the electrical charge pattern that corresponds to the light image falling on the signal electrode.

The photoconductor has a fairly high resistance when in the dark Light falling on the material excites additional electrons into a conducting state, lowering the resistance of the photoconductive material at the point of illumination (*see* PHOTOCONDUCTIVITY). A positive voltage is applied to one side of the photoconductive layer by means of the signal electrode. On the other side the scanning beam deposits sufficient electrons at low velocity to maintain a zero voltage. In the interval between successive scans of a particular spot, the light lowers the resistance in relation to its intensity Current then flows through the surface at this point and the back surface builds up a positive voltage until the beam returns to scan the point. The signal output current is generated when the beam returns this positively charged area to zero voltage. An equal number of electrons flow out of the signal electrode and through a load resistor, developing a signal voltage that is fed directly to a low-noise video-signal amplifier.

A fine-mesh screen stretched across the tube near the target causes the electron scanning beam to decelerate uniformly at all points and approach the target in a perpendicular manner The beam is brought to a sharp focus on the target by the longitudinal magnetic field of the focusing coil and

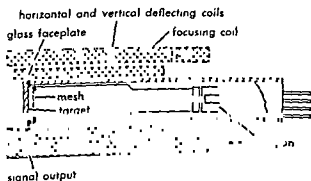


Fig. 7. Cross section of a vidicon tube and its associated deflection and focusing coils.

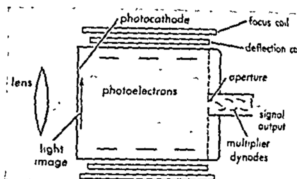


Fig. 8. The image dissector and its associated deflecting and focusing coils. (From D. G. Fink, ed, Television Engineering Handbook, McGraw-Hill, 1957)

the proper voltage for the focusing electrode. The beam is made to scan the target by varying the magnetic fields of the deflecting coils (*see* CATHODE-RAY TUBE). The photoconductor is chosen to have a low secondary-emission ratio and as a result does not charge positively when subjected to electron bombardment. The material that has found the widest use for the vidicon is a form of antimony trisulfide. Ultraviolet-sensitive vidicon tubes employ a thin faceplate of ultraviolet-transmitting glass and a selenium photoconductor that is sensitive to ultraviolet light. An x-ray-sensitive vidicon utilizing a lead oxide photoconductive material has been made for medical and industrial x-ray work. The target is large enough to cover substantial portions of the human anatomy. The scanning beam is focused and deflected by electrostatic electrodes.

Image dissector. The image dissector is a nonstorage type of camera tube that employs a photoemissive light-sensitive surface. This tube does not utilize a scanning beam as do all other types of camera tubes but collects and directly amplifies the electron currents emitted from the photosensitive surface. The image dissector tube was one of the first camera tubes ever devised and is still used in industrial television systems. It is rather insensitive, requiring strong light for proper operation, but it

side of the tube faceplate (Fig. 8). Electron streams from the illuminated photosurface are refocused on an imaging plane at the other end of the tube by the focusing coil and the electric fields of the electrodes in the tube A plate with a small aperture in it is positioned near the center of this plane. Behind the hole is an electron multiplier. This small hole can intercept the electrons from only a few of the many electron streams. Those electrons that pass through the hole are multiplied to produce a high current. The current is collected by the anode of the multiplier and flows through a load resistor placed in the anode circuit. The voltage that develops across this resistor is used to form a television signal

The magnetic fields produced by the deflecting coils shift all the electron streams horizontally and

vertically across the aperture in a regular television scanning sequence. As a result, the aperture periodically samples or dissects the entire photoelectric image. The electrons that do not pass through the aperture are not utilized or stored. This loss contributes to the low sensitivity or low efficiency of this type of nonstorage camera tube. See TELEVISION CAMERA. [R.G.N.]

Bibliography: D. G. Fink (ed.), *Television Engineering Handbook*, 1957; V. K. Zworykin and E. G. Ramberg, *Photoelectricity and Its Application*, 1949.

Television networks

Transmission means to permit the distribution of the same television program simultaneously to two

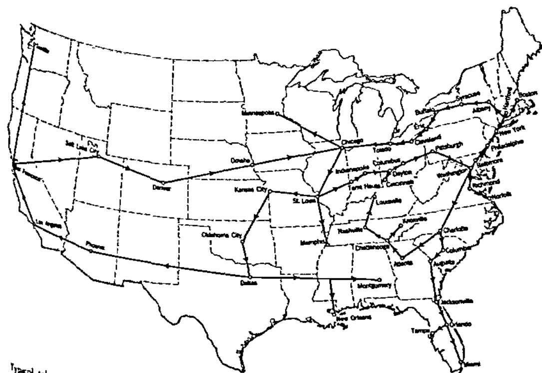
some combination of local cable facilities, microwave radio, and intercity coaxial cable. During the months of daylight saving time some areas continue to operate on standard time, and duplicate facilities may be required in order that programming, which may be repeated live or played back from recordings, may be received at the optimum time. In addition, program material may be recorded for future use in some cities while it is being broadcast in others. See TELEVISION.

Network operation. Much of the network operation is done on a round robin basis. Round robin is network jargon for a circuit in the form of a loop. The station transmitting to a round robin must open the loop at that point so that the program does not circulate around the loop indefinitely. Advantages of round robin operation include the possibility that any station on the loop can originate programs, switching is simplified, and more economical use is made of intercity transmission facilities.

The illustration shows part of a typical network provided by the Bell System for one of the major broadcasting companies. The round robin includes New York, Buffalo, Cleveland, Chicago, St. Louis, Indianapolis, Pittsburgh, Washington, and back to New York. Other cities are served by terminals at intermediate points or by side legs. The main route to the West leaves the round robin at St. Louis, serving intermediate points along the way. Cities in the West are served via Los Angeles, San Francisco, and Salt Lake City. Programs originating on the West Coast also follow these routes.

In network operations, frequent switching of the facilities is required. Switching is done in telephone company television operating centers located in the major cities and other strategic points. Switching can be done on a completely manual basis or on a semiautomatic basis. In semiautomatic operation the switching equipment can be prepared in advance to make one or more simultaneous switches at the operation of a push button. The preparation and operation can be at either the operating center where the circuits terminate or it can be at a remote point. When remote operation is employed, the switching functions are directed over telegraph-type circuits.

The time at which switches are made can be scheduled, or it can be determined on a cue basis.



Typical television network on Bell System facilities.

Most switching is scheduled, but programs with indefinite intervals, such as athletic events, must be switched on cues. Much of the scheduled switching is done 20 seconds before the quarter hour. The interval between switching and the exact quarter hour is usually used by the stations for local commercials or station identification. With all networks being switched at the same time, conflicts are avoided which might be troublesome or embarrassing. This would be particularly true when a facility is scheduled for several networks in sequence. *See* RADIO BROADCASTING NETWORKS.

Technical considerations. Television channels passing through television operating centers are arranged for convenient monitoring and testing. Two monitors are usually provided. An oscilloscope is used to observe the video wave (amplitude versus time) form. The other monitor, essentially a high-grade television receiver, is used to observe picture quality.

Several forms of test signal are employed to determine the condition of the lines. The amplitude versus frequency characteristic can be determined by single-frequency tones in increments from a very low frequency to a frequency at the top of the video band. Precise measurements can be made with a thermocouple-type instrument or, if a reference tone is transmitted simultaneously with the test tones, a comparing bridge can be used. The latter is more accurate since flat changes in gain of the system under test have little effect on the results. Another frequently used test signal is one which sweeps from low to high frequencies. When observed with a suitable detector and oscilloscope, this indicates at a glance the condition of the circuit under test A third test signal which can be used for more gross tests is the multiburst and is observed on the waveform monitor This consists of a square wave pulse followed by several bursts of different frequencies This signal occupies one horizontal video line and may be repeated every line or only on selected lines. In addition to a gross gain versus frequency check it is useful in detecting compression and streaking in the transmission facilities.

Other test equipment measures streaking and smearing, differential gain and phase distortions, noise, and envelope delay distortion.

To reproduce an unimpaired picture, the energy from the television camera should reach the receiver with all frequency components unchanged in relative amplitude and time relationship. No transmission facility is capable of ideal transmission, and therefore the various anomalies must be corrected by electrical networks to approach the ideal as nearly as practicable.

In broadcast television the highest frequency video component transmitted to the receiving sets is approximately 4.2 megacycles (Mc). Ideally, all components from near direct current to 4.2 Mc should be transmitted by the network facility, but economic considerations may dictate a band of lesser width. For example, one of the coaxial-cable carrier systems is capable of transmitting a band approximately 3 Mc wide. This may seem inadequate when viewed on high-quality studio picture monitors, but subjective tests on large numbers of people have shown that the video band can be reduced to below 3 Mc on home television receivers without being noticeable to many.

Local channels. The connections between the intercity television network and the broadcasters' studios are made via local channels which are also used to connect each broadcaster's studio to his transmitter. Local channels are necessarily capable of precise adjustment because of the number which may be in tandem. The distortions, such as amplitude and delay, already present in cameras, studio equipment, long-distance facilities, and the broadcast transmitters, limit the permissible distortion remaining to be allotted among all the local channels in tandem. Coaxial cables are not generally suitable, since the video signals are transmitted at baseband, that is, from essentially direct current to about 4 Mc. Coaxial cables are electrically unbalanced with respect to ground and are susceptible to induced voltages and to differences in potentials at the ends. For this reason local channels in cable are transmitted over balanced video cable pairs A video pair is relatively heavy gage, usually no. 16, and is insulated with a low-loss dielectric. Each video pair is shielded by copper tapes or woven copper wires.

The attenuation of a 16-gage video pair is about 19 db/mile at 4.5 Mc and 75°F. Attenuation varies directly as approximately the square root of frequency; it also varies approximately 0.1% per °F For the degree of precision required in television signal transmission, the temperature variation precludes the use of aerial cable, since the temperature may vary 60°F or more from day to night. Video amplifiers are spaced at intervals up to 4 5 miles to compensate for attenuation.

Capacitance-coupled amplifiers are incapable of transmitting the dc component of video signals Clampers are used to restore the direct current at the end of the circuit One form of clamper samples the deviation of the tip of the sync pulse from its normal level and injects the derived error signal into the video signal in proper phase.

Many local channels are transmitted by microwave radio. The conditions controlling the use of radio or cable are usually economic but may also be influenced by the necessity for a speedy installation, such as for one-time news or sporting events. Where underground cable is not feasible because of terrain, right of way, or other problems, microwave radio becomes attractive even for relatively short distances.

The microwave radio equipment for temporary use is packaged for portability. It is possible, under some conditions, to establish a microwave radio link in less than 1 hour. If, however, tower construction is required for path clearance, it requires more time. Some telephone companies own portable towers, which ca[illegible] erected in a few

tours. Others have made use of truck-mounted cranes as towers for temporary services.

In a heavily congested area, such as New York City the coordination of frequency assignments is a serious problem. For example, the New York Telephone Company has operated as many as 17 microwave radio systems simultaneously on the Empire State Building. Interference between these systems and with long-distance channels in the area is difficult to avoid.

Transmission facilities. The first intercity television network transmission was over a coaxial-cable system known in the Bell System as the L1 carrier system. The television channel uses a vestigial sideband channel transmitting the upper sideband, a 311 kilocycle carrier and a vestigial lower sideband. One form of the coaxial tube has an inside diameter of approximately 0.375 in. and a center conductor approximately 0.10 in. in diameter. The center conductor is supported by polyethylene disks at intervals of about 1 in. The attenuation is approximately 6.75 db/mile at 3 Mc. Repeaters are spaced at about 8-mile intervals. Usually, several coaxial tubes are provided in each direction for service and one tube in each direction for spare. Included also are some conventional telephone circuit conductors for use in various control functions, order wires, and so forth. The useful video bandwidth in an L1 system is slightly less than 3 Mc.

The NTSC (National Television Systems Committee) color television signal contains energy components up to 4.2 Mc. The color information is carried by a subcarrier at about 3.6 Mc. In an L1 carrier system arranged for color transmission the [illegible]

[illegible] the broadcaster. In order to avoid a restricted bandwidth when transmitting monochrome signals on color treated circuits, the carrier terminals are arranged to recognize the presence of the color signal. When the color signal is absent, the L1 carrier terminals transmit the monochrome television signals without the 2-Mc restriction of the video band. *See* COLOR TELEVISION.

A later type of coaxial carrier system, known as the L3, is capable of handling one 4-Mc video channel and 600 message channels in each coaxial tube. Automatic regulation compensates for various changes in the cable and the repeaters. These changes are a function of temperature and time. No special treatment is required to make the L3 carrier system capable of transmitting NTSC color signals.

In both of the L carrier systems, arrangements are made for automatic switching of the circuits to spare facilities in the event of a circuit failure.

Power to the repeaters is supplied by commercial power to main repeaters and over the coaxial cable to auxiliary repeaters. Engine alternator equipment is provided at main repeaters. It automatically starts in the event of interruption of commercial power.

By far the largest number of television circuit miles is provided by microwave radio relay. An example of this is what is known in the Bell System as the TD-2 radio system. This is a frequency-modulated system with six carriers in each direction in a band from about 3700 to 4200 Mc. Radio-relay systems are capable of transmitting video bands of 4.2 Mc or greater. The relay stations are spaced about 30 miles apart.

Microwave radio systems are subject to fading as atmospheric conditions change. Automatic switching equipment is arranged to recognize fading and switch from the faded channel to one of another frequency. As in the L carrier system, emergency power is provided at radio relay stations.

Closed-circuit networks. Closed-circuit television networks are required by industry, government, and educational institutions. They are used for sales promotion, conferences, surveillance, traffic control, and instruction. The same transmission media as that for broadcast television networks may be used or, for distances up to about 35 miles, modulated vhf carriers may be transmitted over coaxial cables. The carriers may be standard broadcast television frequencies between 54 and 88 Mc (channels 2–6) or they may be so-called subchannels between 25 and 50 Mc. This may be particularly attractive economically where more than one channel is required. It is possible to transmit five monochrome channels simultaneously on one coaxial cable in the 54–88-Mc or three in the 25–50-Mc region. By judicious selection of the carrier frequencies it is possible to transmit six channels, three in the 25–50-Mc and three in the 54–88-Mc region. The presence of strong rf signals from TV broadcasting stations or other sources may preclude the use of one or more of the channels if the demodulation equipment is not well shielded.

Experimental work has been done on a wideband video system for closed-circuit systems to transmit high definition color signals. Video signals as wide as 15 Mc have been handled on a limited basis. Special treatment of the network facilities is required when it is necessary to transmit signals of greater bandwidth than the standard broadcast signal. *See* CLOSED-CIRCUIT TELEVISION. [R.B HO.]

Television receiver

The equipment used to receive the transmitted modulated radio-frequency signals and produce synchronized visual images and sound for entertainment or educational purposes. The radio-frequency portion operates on the superheterodyne principle similar to FM and AM receivers. *See* RADIO RECEIVER.

The first television receivers to be mass produced were monochrome, that is, they provided pictures in black and white only. Later, color receivers,

Most switching is scheduled, but programs with indefinite intervals, such as athletic events, must be switched on cues. Much of the scheduled switching is done 20 seconds before the quarter hour. The interval between switching and the exact quarter hour is usually used by the stations for local commercials or station identification. With all networks being switched at the same time, conflicts are avoided which might be troublesome or embarrassing. This would be particularly true when a facility is scheduled for several networks in sequence. *See* RADIO BROADCASTING NETWORKS.

Technical considerations. Television channels passing through television operating centers are arranged for convenient monitoring and testing. Two monitors are usually provided. An oscilloscope is used to observe the video wave (amplitude versus time) form. The other monitor, essentially a high-grade television receiver, is used to observe picture quality.

Several forms of test signal are employed to determine the condition of the lines. The amplitude versus frequency characteristic can be determined by single-frequency tones in increments from a very low frequency to a frequency at the top of the video band. Precise measurements can be made with a thermocouple-type instrument or, if a reference tone is transmitted simultaneously with the test tones, a comparing bridge can be used. The latter is more accurate since flat changes in gain of the system under test have little effect on the results. Another frequently used test signal is one which sweeps from low to high frequencies. When observed with a suitable detector and oscilloscope, this indicates at a glance the condition of the circuit under test. A third test signal which can be used for more gross tests is the multiburst and is observed on the waveform monitor. This consists of a square wave pulse followed by several bursts of different frequencies. This signal occupies one horizontal video line and may be repeated every line or only on selected lines. In addition to a gross gain versus frequency check it is useful in detecting compression and streaking in the transmission facilities.

Other test equipment measures streaking and smearing, differential gain and phase distortions, noise, and envelope delay distortion.

To reproduce an unimpaired picture, the energy from the television camera should reach the receiver with all frequency components unchanged in relative amplitude and time relationship. No transmission facility is capable of ideal transmission, and therefore the various anomalies must be corrected by electrical networks to approach the ideal as nearly as practicable.

In broadcast television the highest frequency video component transmitted to the receiving sets is approximately 4.2 megacycles (Mc). Ideally, all components from near direct current to 4.2 Mc should be transmitted by the network facility, but economic considerations may dictate a band of lesser width. For example, one of the coaxial-cable carrier systems is capable of transmitting a band approximately 3 Mc wide. This may seem inadequate when viewed on high-quality studio picture monitors, but subjective tests on large numbers of people have shown that the video band can be reduced to below 3 Mc on home television receivers without being noticeable to many.

Local channels. The connections between the intercity television network and the broadcasters' studios are made via local channels which are also used to connect each broadcaster's studio to his transmitter. Local channels are necessarily capable of precise adjustment because of the number which may be in tandem. The distortions, such as amplitude and delay, already present in cameras, studio equipment, long-distance facilities, and the broadcast transmitters, limit the permissible distortion remaining to be allotted among all the local channels in tandem. Coaxial cables are not generally suitable, since the video signals are transmitted at baseband, that is, from essentially direct current to about 4 Mc. Coaxial cables are electrically unbalanced with respect to ground and are susceptible to induced voltages and to differences in potentials at the ends. For this reason local channels in cable are transmitted over balanced video cable pairs. A video pair is relatively heavy gage, usually no. 16, and is insulated with a low-loss dielectric. Each video pair is shielded by copper tapes or woven copper wires.

The attenuation of a 16-gage video pair is about 19 db/mile at 4.5 Mc and 75°F. Attenuation varies directly as approximately the square root of frequency; it also varies approximately 0.1% per °F. For the degree of precision required in television signal transmission, the temperature variation precludes the use of aerial cable, since the temperature may vary 60°F or more from day to night. Video amplifiers are spaced at intervals up to 4.5 miles to compensate for attenuation.

Capacitance-coupled amplifiers are incapable of transmitting the dc component of video signals. Clampers are used to restore the direct current at the end of the circuit. One form of clamper samples the deviation of the tip of the sync pulse from its normal level and injects the derived error signal into the video signal in proper phase.

Many local channels are transmitted by microwave radio. The conditions controlling the use of radio or cable are usually economic but may also be influenced by the necessity for a speedy installation, such as for one-time news or sporting events. Where underground cable is not feasible because of terrain, right of way, or other problems, microwave radio becomes attractive even for relatively short distances.

The microwave radio equipment for temporary use is packaged for portability. It is possible, under some conditions, to establish a microwave radio link in less than 1 hour. If, however, tower construction is required for path clearance, it requires more time. Some telephone companies own portable towers, which can be erected in a few

e gain of this amplifier is essentially constant 43 to 45 Mc. Above the latter frequency the onse decreases such that at 45.75 Mc, the picture carrier frequency, it is 50%. This slope is re- ed to compensate for the vestigial sideband smitted signal

elow 43 Mc the response decreases until at 5 Mc, the sound carrier frequency, it is 5–10% he flat response. This minimizes *cross modula-* i between picture and sound carriers. Fixed ed trap circuits are used to produce sharp *cut-s at the lower and upper* limits of the i-f pass- nd. Sufficient selectivity is provided to minimize erference from signals originating in adjacent levision channels.

Separation of video and audio. The output of the f amplifier consists of two modulated rf signals. he of these, which is amplitude modulated, pro- ides a varying signal corresponding to the black nd white *portions of the picture, a blanking signal* o render the return trace invisible on the picture ube, horizontal sync pulses to initiate the retrace of the beam at the end of each line and vertical sync pulses to initiate the retrace of the beam at the end of each picture field. The other signal is frequency modulated and contains the transmitted sound information.

These two rf *signals are applied to a diode,* ei- ther a tube or crystal, which produces a rectified output that follows the instantaneous peak value of the amplitude-modulated picture carrier. The po- larity of this output depends upon the design of the *video amplifier* and method of picture-tube drive. Usually maximum picture carrier (sync- pulse modulation) produces a negative output volt- age.

Coincidentally a 4.5 Mc signal results from the heterodyne beat of the picture and sound carriers This signal contains the frequency-modulated sound information, which can be further amplified and detected *in the sound channel* This is known as the intercarrier sound system (ICS).

drive of the picture tube over its modulation ge. Single-stage pentode amplifiers are gener- r adequate, driving the *picture-tube cathode* h sync positive. A 4.5-Mc trap is included in the eo amplifier *circuit* to prevent the appearance the intercarrier sound signal on the picture tube. For sound reproduction, the intercarrier sound nal formed at the *detector is passed through* a 5-Mc i-f amplifier, some form of amplitude limit- g and to an FM detector, which converts the und carrier modulation to an audio-frequency gnal. This signal is then passed through an audio nplifier to a loudspeaker *and converted* to acous- : output.

Automatic gain control. Since television receiv- s, like radio receivers, may be subjected to widely rying incoming signal strengths, some form of automatic gain control (AGC) is necessary. Cir- cuits for this function provide a nearly constant *carrier signal level to the video* detector by chang- ing the bias on the rf and i-f amplifier tubes as the strength of the incoming signal varies. A simple *form may rectify* the peak amplitude of the de- tected video signal in an *RC* load circuit. Under noisy conditions, however, its performance is not adequate. Improved forms employ an AGC ampli- fier and some type of gating circuit. Time gating

time-gated, and *most noise pulses on the grid* of the tube do not affect AGC action.

Sync separator circuits. Picture synchronizing information *is obtained from the* video signal by means of sync separation circuits. In addition, these circuits must separate this information from noise and interference during the reception of weak signals, particularly if impulse noise is pres- ent. In general, the sync separation circuits per- form the following functions: (1) separation, by means of amplitude clipping, of the sync informa- tion from the picture information; (2) *separation* of the desired horizontal and vertical timing in- formation by means of frequency selection; and (3) rejection *of noise signals that* are higher in amplitude than sync pulses by amplitude limiting or gating (noise suicide) circuits

Sweep systems Two independent sweep systems are employed in the vertical and horizontal sweep circuits. Each employs a timing generator, gener- ally of the oscillatory type, controlled by the syn- chronizing information obtained from the sync sep- arators. The *oscillators are followed by drive and* waveform-shaping circuits These are followed by power amplifier stages capable of providing the currents required by the deflection coils of the yoke for picture-tube beam deflection. Substan- *tially different* techniques are required for vertical and horizontal scanning.

Vertical deflection. Generally the vertical oscil- *lator is of the blocking type operating at approxi-* mately 60 cycles per second (cps). Its frequency is accurately controlled by a signal obtained from the sync separator The output waveform of the sync separator consists of a train of pulses repre- *senting the horizontal* and vertical synchronizing pulses. When these are passed through a low-pass filter or integrating circuit, a saw-tooth-shaped voltage wave *representing vertical sync is obtained.* This is applied to the grid of the vertical oscillator. A frequency control in the vertical oscillator cir- cuit is so adjusted that its free-running frequency is slightly lower than the synchronizing signal *frequency. For good interlace it is necessary* that no horizontal frequency components be included in the vertical synchronizing voltage.

The vertical output stage is generally *operated* as a class A amplifier (*see* POWER AMPLIFIER). The *yoke is transformer-coupled to the plate* of the out- put tube to match the yoke impedance to the output

which produce pictures in full color or in black and white, became available. For basic discussion of a television system, *see* TELEVISION; TELEVISION STANDARDS.

Monochrome receivers. Figure 1 shows a block diagram of a conventional monochrome television receiver, the major sections of which will be discussed in the following paragraphs.

Antenna and transmission line. Since all broadcast television transmissions in the United States are horizontally polarized, the most basic type of television-receiving antenna is the horizontally mounted half-wave dipole. Because the stations serving a given area may operate on widely different frequencies, however, the dipole dimensions must be a compromise that permits reasonable performance on all the desired channels. *See* ANTENNA (AERIAL). More complex antennas combine several dipole elements of various lengths, and passive reflectors may be used to achieve some degree of horizontal directivity. Highly directive antennas are frequently mounted on remotely controlled rotators so they can be pointed in the direction providing the best reception of the desired signal. The most common types of transmission line between the antenna and receiver are 300-ohm "twin-lead," employing polyethylene as a dielectric spacer between two uniformly spaced, unshielded wires Also 75-ohm coaxial cable is used. *See* TRANSMISSION LINES.

Tuner. The tuner of a television receiver selects the desired channel and converts the frequencies received to lower frequencies within the pass-band of the intermediate-frequency amplifier. For very-high-frequency (vhf) reception the tuner generally has 12 discrete positions, corresponding to channels 2–13. For ultra-high-frequency (uhf) reception, continuous tuning is employed. Nearly all vhf tuners employ radio-frequency (rf) amplifier, mixer, and local-oscillator circuits arranged as shown in Fig. 1. In uhf tuners the rf amplifier is sometimes omitted because of the difficulty of obtaining low-noise amplification at uhf frequencies. The rf amplifier may be of the cascode, tetrode, or pentode type. In general, the cascode circuit provides superior results. *See* CASCODE AMPLIFIER.

The mixer and local-oscillator circuits may employ separate tube envelopes or be combined in the same glass envelope. The received signal and the local oscillator signal are applied to the mixer. Difference frequencies, representing the picture and sound carriers, are produced and remain essentially constant as the rf amplifier, mixer, and oscillator circuits are tuned to the different channels. Known as intermediate frequencies (41.25 Mc for sound and 45.75 Mc for picture), they are available for further amplification. The correct oscillator frequency is approximately set at the time of channel selection. A fine adjustment is provided to permit more accurate tuning.

Such performance characteristics as noise factor, gain, bandwidth, and oscillator radiation must be optimized in the design of the tuner.

Intermediate-frequency amplifier. The output from the tuner is applied to the intermediate-frequency (i-f) amplifier. Several stages of amplification are required to obtain the desired output signal level and selectivity.

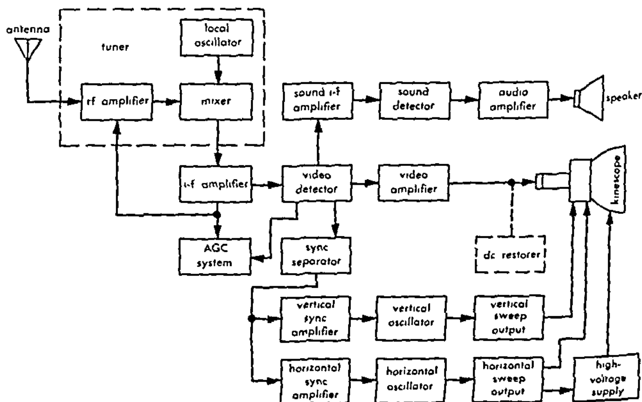


Fig. 1. Block diagram of a typical monochrome television receiver

The gain of this amplifier is essentially constant from 43 to 45 Mc. Above the latter frequency the response *decreases* such that at 45.75 Mc, the picture carrier frequency, it is 50%. This slope is required to compensate for the vestigial sideband transmitted signal.

Below 43 Mc the response decreases until at 41.25 Mc, the sound carrier frequency, it is 5–10% of the flat response. This minimizes cross modulation between *picture and sound carriers*. Fixed tuned trap circuits are used to produce sharp cutoffs at the lower and upper limits of the i-f passband. Sufficient selectivity is provided to minimize interference from signals originating in adjacent television channels.

vides a varying signal corresponding to the black and white portions of the picture, a blanking signal to render the return trace invisible on the picture tube, horizontal sync pulses to initiate the retrace of the beam at the end of each line and vertical sync pulses to initiate the retrace of the beam at the end of each picture field. The other signal is frequency *modulated and contains the transmitted* sound information.

These two rf signals are applied to a diode, either a tube or crystal, which produces a rectified output that follows the instantaneous peak value of the amplitude-modulated picture carrier. The polarity *of this output* depends upon the design of the video *amplifier and method of picture-tube* drive. Usually maximum picture carrier (sync-pulse modulation) produces a negative output voltage.

Coincidentally a 4.5 Mc signal results from the heterodyne beat of the picture and sound carriers. This signal contains the frequency-modulated sound information, which *can be further amplified* and detected in the sound channel. This is known as the intercarrier sound system (ICS).

Following the detector is a video amplifier, which consists of one or two stages depending upon the over all receiver requirements. An output level of about 100 volts is ordinarily sufficient to assure full drive of the picture tube *over its modulation* range. Single-stage pentode amplifiers are generally adequate, driving the picture-tube cathode with sync positive. A 4.5-Mc trap is included in the video amplifier circuit to prevent the appearance of the intercarrier sound signal on the picture tube.

For sound reproduction, the intercarrier sound signal formed at the detector is passed *through* a 4.5-Mc i-f *amplifier*, some form of amplitude limiting, and to an FM detector, which converts the sound carrier modulation to an audio-frequency signal. This signal is then passed through an audio amplifier to a loudspeaker and converted to acoustic output.

Automatic gain control. Since television receivers, like radio receivers, may be subjected to widely varying incoming signal strengths, some form of automatic gain control (AGC) is necessary. Circuits for this function provide a nearly constant carrier signal level to the video detector by changing the bias on the rf and i-f amplifier tubes as the strength of the incoming signal varies. A simple form may *rectify the peak amplitude* of the detected video signal in an *RC* load circuit. Under *noisy conditions*, however, its performance is not adequate. Improved forms employ an AGC amplifier and some type of gating circuit. *Time gating* voltage is derived from the horizontal sweep circuit during the beam scanning return time. In this way the plate current of the AGC amplifier tube is time-gated, and most noise pulses on the grid of the tube do not *affect AGC action*.

Sync separator circuits. Picture synchronizing *information* is obtained from the video signal by means of sync separation circuits. In addition, these circuits must separate this information from noise and interference during the reception of weak signals, particularly if impulse noise is present. In general, the sync separation circuits perform the following functions: (1) separation, by means of *amplitude clipping, of the sync information* from the picture information; (2) separation of the desired horizontal and vertical timing information by means of frequency selection; and (3) rejection of noise signals that are higher in amplitude than sync pulses by amplitude limiting or gating (noise suicide) circuits.

Sweep systems. Two independent sweep systems are employed in the vertical and horizontal sweep *circuits. Each employs a timing generator*, generally of the oscillatory type, controlled by the synchronizing information obtained from the sync separators. The oscillators are followed by drive and waveform-shaping circuits. These are followed by power amplifier stages capable of providing the currents required by the deflection coils of the yoke *for picture-tube beam deflection*. Substantially different techniques are required for vertical and horizontal scanning.

Vertical deflection. Generally the vertical oscillator is of the blocking type operating at approximately 60 cycles per second (cps). Its frequency is accurately controlled by a signal obtained from *the sync separator*. The output waveform of the sync separator consists of a train of pulses representing the horizontal and vertical *synchronizing* pulses. When these are passed through a low-pass filter or integrating circuit, a saw-tooth-shaped voltage wave representing vertical sync is obtained. This is applied to the grid of the vertical oscillator. *A frequency control* in the vertical oscillator *circuit* is so adjusted that its free-running frequency is slightly lower than the *synchronizing signal* frequency. For good interlace it is necessary that no horizontal frequency components be included in the vertical synchronizing voltage.

The vertical output stage is generally operated as *a class A amplifier (see* POWER AMPLIFIER). The yoke is transformer-coupled to the plate of the output tube to match the yoke impedance *to the output*

tube impedance. Since the yoke impedance is partly resistive and partly inductive, the voltage waveform across it is the sum of a saw-tooth and a rectangular pulse. The current through the yoke has essentially a saw-tooth waveform, but each saw-tooth has a symmetrical S shape to take care of picture-tube face-plate geometry and result in a linear scan.

Horizontal deflection. A more complex system is required for horizontal scanning. There are several basic reasons for this: (1) horizontal sync pulses are of much shorter duration than are vertical sync pulses; (2) some form of automatic frequency control (AFC) of the horizontal oscillator is required to average the incoming horizontal sync information and retain accurate phase; and (3) considerably greater power output is required to generate the deflecting yoke fields as well as the high voltage (10–20 kilovolts) for the picture tube.

The horizontal oscillator is generally of the blocking type. The frequency of oscillation is determined both by a time-constant control and by a bias voltage derived from an AFC circuit. The AFC circuit may be a phase comparator, in which the pulses from the sync separator are compared to the oscillator output signal The output of the comparator is a voltage proportional to the phase departure of the two signals.

The desired current waveform in the horizontal windings of the deflection yoke is a line-frequency saw-tooth, possibly modified by the addition of a small amount of S curvature to compensate for picture-tube face geometry. Energy from the horizontal driver, or output tube, is normally supplied to the yoke (through the horizontal output transformer) only during approximately the last half of each saw-tooth period. At the conclusion of the saw-tooth period, the horizontal driver is cut off, and the energy stored in the form of current through the yoke causes an oscillation in the self-resonant circuit consisting of the yoke, horizontal output transformer, and the associated capacitances. This oscillation is permitted to continue for only one half-cycle, during which time the current through the yoke reverses in polarity and attains a negative value almost equal to the original positive value. The self-resonant frequency of the horizontal output circuit must be high enough to permit the full current reversal to be accomplished within the horizontal blanking interval. The oscillation is stopped after the first half-cycle by the action of a damper tube (normally a diode), which controls the release of the energy stored in the yoke in such a way that the current follows the desired saw-tooth waveform. In approximately the middle of the saw-tooth period, the damper tube becomes nonconductive, and the horizontal driver tube takes over the task of supplying the energy required for the next cycle.

High-voltage supply. Since the impedance of the yoke at horizontal scan frequency is primarily inductive, the voltage across the horizontal deflection windings is essentially constant during active scan. During the retrace period, however, the high rate of current change causes the generation of a high voltage pulse having a shape similar to a half sine wave and a duration equal to the retrace period. It is common practice to employ a step-up winding on the horizontal output transformer to raise this so called kickback pulse up to a still higher voltage level, commonly about 18 kilovolts (kv) and to pass it through a simple rectifier and filter to serve as the high-voltage supply for the kinescope

Picture tubes. The display device for a television receiver is a cathode-ray tube, consisting of an evacuated bulb containing an electron gun and a phosphor screen, which emits light when excited by an electron beam. The intensity of the electron beam is controlled by the video signal, which is applied either to the grid or the cathode of the electron gun. The position of the electron beam is controlled by electromagnetic fields produced by the deflection yoke placed around the neck of the tube. *See* KINESCOPE.

Controls. Certain controls are available to the user for adjustment of the receiver. These are the audio volume, channel selector, fine tuning, brightness, contrast, horizontal hold, and vertical hold controls. Other controls, normally mounted on the rear of the chassis or under a removable panel, include height, width, and linearity controls.

The ON-OFF switch for the receiver is frequently mounted on the same shaft as the audio volume control, which controls the gain of the audio channel. The channel selector adjusts the tuner's selective circuits for optimum performance at the desired channel, and fine tuning is a vernier control for the frequency of the local oscillator. Brightness is usually a manual adjustment of the bias on the electron gun in the picture tube. The contrast control adjusts the level of the video signal, by some such means as a variable resistor in the cathode circuit of one of the video amplifier stages.

The horizontal and vertical hold controls adjust the free-running frequencies of the horizontal and vertical oscillators to achieve the most reliable synchronization with the incoming signal. In both cases, the controls may actually consist of variable resistors in the grid circuits of the respective blocking oscillators.

Vertical linearity is generally controlled by a variable resistance in the grid circuit of the vertical output stage, and picture height may be controlled by a variable resistor in the plate circuit of the vertical blocking oscillator. The width control may be a variable resistor in the screen grid circuit of the horizontal output tube, and horizontal linearity may be controlled by a variable inductor placed between the damper tube plate and the source of plate voltage.

Color receivers. Television receivers designed to produce images in full color are necessarily more complex than those designed to produce monochrome images only, because additional information must be handled to produce color. In mono-

chrome systems, the video signal controls only the luminance of the various areas of the image. In color systems, it is necessary to control both the luminance and chrominance of the picture elements.

The chrominance of a color refers to those attributes which cause it to differ from a neutral (white or gray) color of the same luminance While chrominance can be expressed in a great variety of ways, it is always *necessary to employ* at least two variables to express the full range of chrominance that can be perceived by the human eye. In qualitative terms, chrominance may be regarded as those properties of a color that control the *psychological sensations of hue and saturation.* For color television purposes, chrominance is most frequently expressed quantitatively in terms of the amounts of two hypothetical, zero-luminance primary colors (usually designated I and Q), which *must be added to or subtracted from a neutral color* of a given luminance to produce the color in question.

As a practical matter, color television receivers produce full-color images as additive combinations of red, green, and blue primary-color images, and it is necessary to process the luminance and chrominance information contained in a color signal in such a way as to make it usable by a practical reproducing device.

Nature of the color signal Color television ...adcasts in the United States employ signal spec...ations that are fully compatible with those used ... monochrome, *making it possible for color pro...ams* to be received on monochrome receivers and ...nochrome programs to be received on color re...ivers. (Color pictures are produced, of course, ...ly when color programs are viewed through color ...ceivers—in all other cases the images are in ...ack and white only.) Compatibility is achieved ... encoding the color information at the transmit...ng end of a color television system in such a way ...at the transmitted signal consists essentially of ... normal monochrome signal (conveying lumi...ance information) supplemented by an additional ...odulated wave conveying chrominance informa...ion. Figure 2 shows the *major components of a* ...olor television signal. Although it is added di...rectly to the monochrome signal component before ...transmission, the color subcarrier signal does not ...ause objectionable interference, because of the ...se of the frequency interlace technique Because ...the chrominance information involves two vari...ables, the modulated subcarrier signal varies in ...both amplitude and phase, and it is necessary to ...employ synchronous detectors to recover the two ...variables. A phase reference for the special local ...

...samples of unmodulated subcarrier trans...mitted during the horizontal blanking periods ...after the horizontal sync pulses. *See* COLOR TELE...VISION.

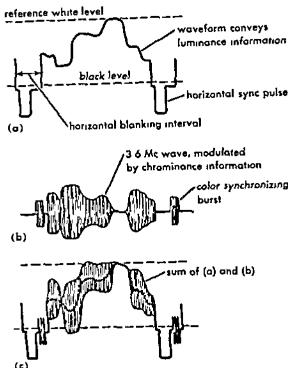


Fig 2 Waveform sketches showing the major components of a compatible color television signal (a) Normal monochrome signal (b) Color subcarrier signal (c) Complete color signal.

Over-all color receiver A simplified block diagram for a color television receiver is shown in Fig. 3 Many of the circuits in a color receiver are the same in principle as the corresponding circuits in a monochrome receiver, so it is unnecessary to redescribe them in detail. It is important to recognize, however, that all circuits handling the complete color signal must be designed for high performance standards. Because the chrominance information is received in the form of sidebands occupying the upper portion of the video spectrum (centered on approximately 3.6 Mc), it is necessary that the antenna, tuner, i-f amplifier, and video detector be designed to handle the full 4-Mc bandwidth provided in the broadcast transmission standards if degradation of the color information is to be avoided. Because the color subcarrier signal is simply added to the normal monochrome signal before transmission, it is necessary that all stages handling the complete signal be linear, so as to avoid intermodulation or distortion of the various signal components The deflection circuits for a color receiver are similar in principle to those used in monochrome receivers, although the output stages are normally designed for a higher power level because of the greater deflection requirements for color kinescopes.

Color decoding circuits Special decoding circuits are necessary in a color receiver to process the luminance and chrominance information in a color signal so that it can be used for the control of a practical color kinescope utilizing red, green,

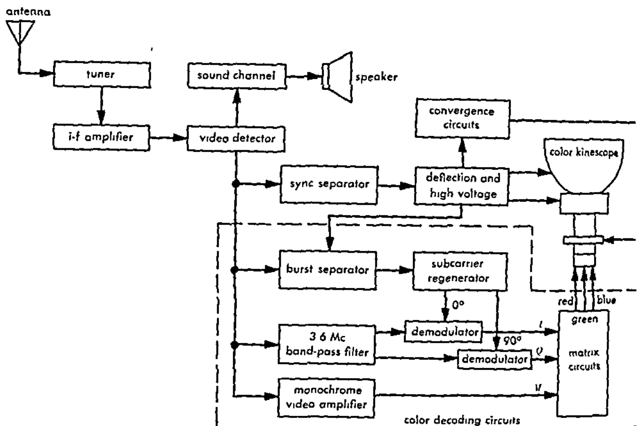


Fig. 3 Simplified block diagram of a color television receiver

and blue primary colors. The major features of the most common approach to color decoding circuits are shown within the dotted lines in Fig. 3

The video amplifier shown at the bottom handles the monochrome portion of the signal and is designed to provide attenuation in the vicinity of 3.6 Mc to block the passage of chrominance information. The chrominance information is recovered from the modulated subcarrier signal through a bandpass filter (centered at 3.6 Mc) and a pair of synchronous demodulators, in which the modulated wave is heterodyned against fixed carriers of two different phases but of the same frequency. In the most rigorous type of color decoding circuit, the chrominance components recovered from the modulated subcarrier signal are the same I and Q originally used to produce the modulated wave, but it is possible to use almost any two phase positions (not necessarily 90° apart) to recover any two independent combinations of the original I and Q signals. The bandwidths of the signals produced by the demodulators are normally adjusted somewhere between 0.5 and 1.5 Mc, and delay compensation may be required to keep all three signal components in time coincidence. The matrix circuit is essentially a linear cross-mixing network for combining the M, I, and Q signals in the proper proportions to produce red, green, and blue signals. If signals other than I and Q are produced by the chrominance demodulators, it is necessary only to design the matrix circuit with slightly different mixing constants.

The synchronous carriers required for the demodulation of the chrominance information are provided by a subcarrier regenerator, which is usually a burst-controlled oscillator operating at the subcarrier frequency. Control information for the subcarrier regenerator is obtained from a burst separator, which is a gate circuit turned on only during the horizontal blanking periods by pulses derived from the horizontal deflection system. The separated bursts are compared with the output of the local subcarrier oscillator in a phase detector If an error exists, a correction voltage is developed, which may be applied through a reactance tube to restore the subcarrier oscillator to the proper frequency and phase. For good noise immunity, a time constant is normally provided so that control information is averaged over at least several line periods.

Color kinescope and convergence circuits. The great majority of color television receivers employ the shadow-mask color kinescope, in which color images are produced in the form of closely intermingled red, green, and blue dots. The primary-color phosphor dots are excited by three separate electron beams, which are prevented from striking dots of the wrong color by the shadowing effect of an aperture mask located about ½ in. behind the special phosphor screen. The three beams in such a kinescope are all deflected simultaneously by the fields produced by a single deflection yoke placed in the conventional position around the neck of the tube. It is necessary, however, to provide an auxiliary deflection system to maintain convergence of the three beams in all areas of the viewing screen, so that the primary-color images are properly registered. A special convergence yoke is commonly

ced around the neck of the kinescope, just ad of the electron guns; this yoke has separate ls for electromagnetic control of the positions of three beams Appropriate waveforms for the vergence yoke coils are derived from the basic lection waveforms In general, each converce coil requires a different combination of sawth and parabolic waveforms at the horizontal l vertical scanning frequencies.

In other types of color kinescopes, not yet develed commercially, the need for convergence circts may be eliminated by the use of a single-gun proach Such alternative kinescopes may have her special requirements, such as precise posin control of the single beam or special gating r the video signal.

Controls. In addition to the same controls reured for monochrome receivers, color receivers rmally have controls for convergence, hue, and turation. The convergence controls, considered rvicing adjustments only, adjust the relative amlitudes and phases of the signal components that

nd alters all the colors in the image in a systemtic manner comparable to the effect achieved hen a color circle diagram is rotated in one direcion or the other. The proper setting for the hue ontrol is normally determined by observing skin ones on actors and actresses. The saturation control, frequently labeled chroma or simply color, adjusts the gain of the chrominance circuits relative to the monochrome channel and controls the saturation or vividness of the reproduced colors. When this control is set too low, the colors are all pale or pastel, and when it is reduced to zero, the picture is seen in black and white only. [C.M.S., J.W.WN.]

Bibliography: S. Deutsch, *Theory and Design of Television Receivers*, 1951; D. G. Fink, *Television Engineering*, 2d ed., 1952, J. W. Wentworth, *Color Television Engineering*, 1955.

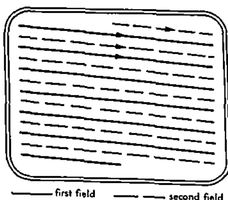
Television scanning

The process of scrutinizing the brightness of each element of detail contained in the image of a scene to be transmitted by television. In monochrome television, the process is instrumental in converting the brightness of each individual element so scrutinized into a unique voltage-time response suitable for transmission. In color television, the brightness variations of each scene are first separated by red, green, and blue filters, after which the conversion from brightness to voltage on a time basis occurs separately for each of the three colors Scanning also takes place in the receiver in exact synchronism with the camera tube scanning, and synchronizing signals are transmitted for that purpose (see TELEVISION).

Interlaced scanning. An image is analyzed by scanning it according to a fine structure of parallel, nearly horizontal lines called a scanning raster The complete raster is rectangular in shape. Scanning may be done conventionally by starting at the upper left-hand corner along line 1 and moving toward the right at constant speed. At the end of line 1 a quick return is made to the left-hand side to start the scanning of line 2, again moving toward the right. When all lines have been scanned in this way, from top to bottom, the process is repeated by returning quickly to the upper left-hand corner to line 1. If all the lines are scanned in sequence, the process is called sequential scanning.

A variation of this kind of scanning, called interlaced scanning, is used in television practice to conserve bandwidth in the transmission system without introducing intolerable flicker. Flicker is a function of the frequency of repetition of coverage of the raster. With interlaced scanning alternate (odd-numbered) lines are scanned first, and the remaining (even-numbered) lines are scanned next. The entire raster area is covered or scanned twice. Therefore, the picture repetition rate for interlaced scanning is twice that of sequential scanning, which is at the same velocity along each line, and a corresponding reduction in the sensation of flicker is obtained. This is called double interlacing and is standard in all broadcast television systems. The entire raster is covered 30 times per second, therefore, the picture area is scanned 60 times a second.

Lines and frequency of scanning. The standards require that 15,750 lines be scanned per second With a vertical scanning rate of 30 times a second, there are 525 lines allocated to each frame and 262.5 lines to each field. Each line is 63.49 microseconds (μsec) in duration. A finite period of time is required to return the scanning beam of electrons in the camera or picture tube to the left edge of the scene for the next line. This period, blanking or retrace time, requires 16–18% of the total line time or 10.16–11.43 μsec. Similarly, 7.5–8% of the vertical field period, or 1250–1333 μsec, is required for the scanning beam to return to the top of the picture. This is the equivalent of about 21 lines of time Blanking circuits prevent the transmission of brightness variations during



The scanning sequence for interlaced scanning The spacing between the lines is enormously exaggerated. (From F E Terman, *Electronic and Radio Engineering*, McGraw-Hill, 1955)

both the horizontal and vertical retrace intervals.

The scanning process is shown in the illustration. Line 1 begins at the top center of the image. The beam proceeds for half a line to the right edge. Retrace to the left occurs, and line 3 follows. At the right end of line 3, one and a half lines have been scanned. Line 5 and successive odd-numbered lines are scanned, for a total of 241.5 lines, ending at the lower right corner. Twenty-one lines elapse during the vertical retrace interval while the scanning beam is moved to the upper left corner, placing it in position to start scanning line 2. This completes one full field. Then, 241.5 successive even-numbered lines are scanned, ending at the middle bottom of the scene. Again 21 lines elapse for vertical retrace, during which time the scanning beam is returned to the top center of the scene. This completes the second field and a full frame. The sequence then repeats.

Resolution. The resolution of a television system is a measure of its ability to reproduce fine detail. It is measured in terms of the number of lines appearing in the reproduced image of a test pattern at the output of a system. The diameter of the electron scanning beam in the camera tube limits the resolution. The smaller the spot, the higher the resolution. The frequency bandwidth f also limits resolution in the following manner

$$f = 0.0125\,n$$

where n is the number of lines and f is in megacycles (Mc). The resolution is 320 lines when the bandwidth is 4.0 Mc. A camera tube may have a resolution of 800 lines, but bandwidth limitations, notably in the transmitter, will reduce that figure to about 300. *See* TELEVISION STANDARDS.

Color coding. A color television camera has separate tubes for red, green, and blue, each of which scans in the same manner as for monochrome. The three camera signals are combined in a colorplexer into a single signal containing both luminance and chrominance information. *See* COLOR TELEVISION. [R.C.K.]

Television standards

Numerical and graphical criteria which describe essential aspects of a television system, employed in the design and operation of equipment to assure that the various parts of the system (studios, transmitters, networks, and receivers) will operate in cooperative fashion at maximum performance. Television systems have a special need, compared with other communication systems, for definitive standards because television transmitters and receivers must operate in a precise "lock-and-key" relationship. In particular, the scanning of the image in the camera must be matched by the scanning in every associated receiver within a timing precision of approximately one-tenth of a millionth of a second, and with relative positions of picture details correct to a few thousandths of an inch.

To assure that any television receiver can receive programs from any transmitter within range, it is customary to set up a *single* set of standard within a group of *neighboring countries*. The systems of Mexico, the United States, Canada, an Cuba operate on one such set of standards. Tho of Europe (except in England and France and i certain transmissions in Belgium) also conform t a single set of standards, which differ in deta from those used in North America. Televisio standards are promulgated, and conformity to the by television stations is legally enforced, by governmental agencies having responsibility for communications systems in the respective countrie In the United States the responsible agency is th Federal Communications Commission. Conformit is not usually enforced in the case of television receiver design, but is assured by the desire of designers to make best use of available signals.

Television standards are of two kinds. Transmission standards describe the signals radiated b television transmitters and their relation to th intended visual and aural content of the program Allocation standards define the manner in whic the radiated signals occupy the radio spectrum These affect primarily the distances over whic television stations provide satisfactory service i the presence of interference from transmitters, receivers, and other sources.

Transmission standards. These standard primarily affect the quality of picture reproduction. They include scanning, modulation, and channel standards.

Scanning standards. The limit of pictorial excellence is established by scanning standards which define the number of lines into which the image is dissected in the scanning process, the pattern in which the scanning lines are laid down, the number of picture scans completed per second, the directions in which scanning proceeds, and the ratio of the picture width to its height.

The number of lines must be sufficient to provide a fine enough picture structure to satisfy the eye at normal viewing distances. Values of 405, 525, 625, and 819 lines per picture are standardized in various countries. The standard in North America is 525 lines. This number is the total of all the lateral traverses in the scanning pattern from the beginning of one picture to the beginning of the next.

Only about 90% of the total number of lines is actually visible in the reproduced image, since some are blanked out in the interval between picture scans and others are hidden by the mask which frames the picture tube. The horizontal scanning direction, standardized throughout the world, is from left to right along each line, and the vertical direction is from the top to the bottom of the picture.

The number of pictures scanned per second is chosen to assure the illusion of smooth motion. A further consideration, to simplify receiver design, is that this number should be one-half the alternating-current frequency of the electric power system. In North America, 60-cycle power predominates, and the number of pictures per second is

standardized at 30; in Europe, 50-cycle power is used and the standard is 25 pictures per second. Higher numbers of pictures per second are not chosen, since this would require more space in the radio spectrum *for each station* for otherwise equally excellent picture quality.

When the pictures are scanned at 25 or 30 per second, flicker is evident unless precautions are taken. To avoid flicker, the viewing screen is illuminated twice during each picture scan. This is accomplished by interlaced scanning. *See* TELEVISION SCANNING.

The aspect ratio is the ratio of the picture width to its height, as scanned within the camera. This ratio must be matched by receivers if the reproduced objects are to have the correct relative

[illegible]

cent use of standard motion picture film which, prior to the advent of widescreen techniques, was defined as 4 by 3 units.

Modulation standards. This second group of transmission standards defines the method by which the sound and picture information is carried by the respective radio carrier signals. Frequency modulation is used for the sound transmission (except in England, France, and in some transmission in Belgium, where amplitude modulation is used). The sound standards define the maximum deviation of the carrier frequency and the amounts of preemphasis given to the higher audio frequencies. The picture information is universally transmitted

[illegible]

ulation standards is the maximum video sideband frequency which is radiated by the transmitter. This value determines the fineness of the picture structure along each line (horizontally) and this must match approximately the fineness of the vertical structure determined by the number of scanning lines. The maximum video frequency is standardized at 3 megacycles (Mc) for the British 405-line system, 4 Mc for the American 525-line

[illegible]

The polarity of picture modulation must also be standardized. In negative modulation, used in the American and European systems, the power of the transmitter decreases as the scene becomes brighter. Positive modulation (power and brightness increase together) is used in England, France, and partially in Belgium. Receiver designers must fit the polarity standard of their region, otherwise the pictures would be reproduced with tonal values reversed, as in a photographic negative. To maximize coverage for given transmitter power, the power level corresponding to full black (absence of light) in the scene is maintained constant. This standard is known as dc transmission and is used throughout the world.

pa

to allow the [illegible]

with minimum interference within the transmission channel. The subcarrier sideband frequency carrying the color information is standardized relative to the picture carrier frequency to a precision of better than 1 part in 3,000,000. The number of lines per picture and pictures per second are, in turn, rigidly related to this subcarrier frequency. The color modulation standards specify the relationship between the relative intensities of the three primary colors and the reference white, and the corresponding amplitude and phase modulations of the color subcarrier. The chromaticities (color coordinates specifying hue and chroma) of the three primary colors and the reference white of the system are also standardized, as are the frequency limits associated with the subcarrier modulation components.

Channel standards. This final group of transmission standards specifies the frequency limits of the station channels and the manner in which the transmitted signals occupy the channels. The channel width must accommodate the maximum video sideband frequency as one sideband, plus a vestigial portion for the other sideband, plus room for the sound transmission and guard bands to prevent interference. Accordingly, the total channel width is 5 Mc for the 405-line British system, 6 Mc for the 525-line American system, and 7 Mc for the 625-line European system. The French system, which provides additional space for video modulation and guard bands, uses a 14-Mc channel.

Allocation standards. These standards define minimum and maximum transmitter powers in relation to antenna tower heights as well as required or permissible signal strengths at specified distances. Also specified are the minimum separation in miles and the minimum frequency separation between channels among stations which serve a given region or which are otherwise capable of causing mutual interference. In the United States, to avoid interference between receivers (the local oscillators of which can radiate interfering signals through the receiving antenna), standard values of intermediate frequency and local oscillator frequency are assumed in the allocation plan. These are closely adhered to by receiver designers. In the United States, the minimum effective radiated transmitter power is specified from 1–50 kw depending on the population to be served. Maximum powers from 100–5,000 kw are specified, depending on the channel used, the height of the antenna tower, and the geographical region in which the station is located.

Equipment standards, to assure compliance with the transmission and allocation standards, are also set up by government regulation and have been adopted, with suitable safety factors, by the trade associations. [D.G.F.]

Bibliography: D. G. Fink (ed.), *Television Engineering Handbook*, 1957.

both the horizontal and vertical retrace intervals.

The scanning process is shown in the illustration. Line 1 begins at the top center of the image. The beam proceeds for half a line to the right edge. Retrace to the left occurs, and line 3 follows. At the right end of line 3, one and a half lines have been scanned. Line 5 and successive odd-numbered lines are scanned, for a total of 241.5 lines, ending at the lower right corner. Twenty-one lines elapse during the vertical retrace interval while the scanning beam is moved to the upper left corner, placing it in position to start scanning line 2. This completes one full field. Then, 241.5 successive even-numbered lines are scanned, ending at the middle bottom of the scene. Again 21 lines elapse for vertical retrace, during which time the scanning beam is returned to the top center of the scene. This completes the second field and a full frame. The sequence then repeats.

Resolution. The resolution of a television system is a measure of its ability to reproduce fine detail. It is measured in terms of the number of lines appearing in the reproduced image of a test pattern at the output of a system. The diameter of the electron scanning beam in the camera tube limits the resolution. The smaller the spot, the higher the resolution The frequency bandwidth f also limits resolution in the following manner.

$$f = 0.0125\,n$$

*where n is the number of lines and f is in mega*cycles (Mc). The resolution is 320 lines when the bandwidth is 4.0 Mc. A camera tube may have a resolution of 800 lines, but bandwidth limitations, notably in the transmitter, will reduce that figure *to about 300 See* TELEVISION STANDARDS.

Color coding. A color television camera has separate tubes for red, green, and blue, each of which scans in the same manner as for monochrome The three camera signals are combined in *a colorplexer into a single signal containing both* luminance and chrominance information *See* COLOR TELEVISION. [R C K.]

Television standards

Numerical and graphical criteria which describe essential aspects of a television system, employed in the design and operation of equipment to assure that the various parts of the system (studios, transmitters, networks, and receivers) will operate *in cooperative* fashion at maximum performance. Television systems have a special need, compared with other communication systems, for definitive standards because television transmitters and receivers must operate in a precise "lock-and-key" *relationship*. In particular, the scanning of the image in the camera must be matched by the scanning in every associated receiver within a timing precision of approximately one-tenth of a millionth of a second, and with relative positions of picture details correct to a few thousandths of an inch.

To assure that any television receiver can receive programs from any transmitter within range, it is customary to set up a single set of standards within a group of neighboring countries. The systems of Mexico, the United States, Canada, and Cuba operate on one such set of standards. Those of Europe (except in England and France and in certain transmissions in Belgium) also conform to a single set of standards, which differ in detail from those used in North America. Television standards are promulgated, and conformity to them by television stations is legally enforced, by governmental agencies having responsibility for communications systems in the respective countries. In the United States the responsible agency is the Federal Communications Commission. Conformity is not usually enforced in the case of television receiver design, but is assured by the desire of designers to make best use of available signals.

Television standards are of two kinds. Transmission standards describe the signals radiated by television transmitters and their relation to the intended visual and aural content of the program. Allocation standards define the manner in which the radiated signals occupy the radio spectrum. These affect primarily the distances over which television stations provide satisfactory service in *the presence of interference* from transmitters, receivers, and other sources.

Transmission standards. These standards primarily affect the quality of picture reproduction. They include scanning, modulation, and channel standards.

Scanning standards. The limit of pictorial excellence is established by scanning standards which define the number of lines into which the image is dissected in the scanning process, the pattern in *which the scanning lines are* laid down, the number of picture scans completed per second, the directions in which scanning proceeds, and the ratio of the picture width to its height.

The number of lines must be sufficient to provide *a fine enough picture structure* to satisfy the eye at normal viewing distances. Values of 405, 525, 625, and 819 lines per picture are standardized in various countries The standard in North America is 525 lines This number is the total of all the lateral *traverses in the scanning* pattern from the beginning of one picture to the beginning of the next

Only about 90% of the total number of lines is actually visible in the reproduced image, since some are blanked out in the interval between picture scans and others are hidden by the mask which frames the picture tube. The horizontal scanning direction, standardized throughout the world, is from left to right along each line, and the vertical direction is from the top to the bottom of the picture.

The number of pictures scanned per second is chosen to assure the illusion of smooth motion. A further consideration, to simplify receiver design, is that this number should be one-half the alternating-current frequency of the electric power system. In North America, 60-cycle power predominates, and the number of pictures per second is

standardized at 30; in Europe, 50-cycle power is used, and the standard is 25 pictures per second. Higher numbers of pictures per second are not chosen, since this would require more space in the radio spectrum for each station for otherwise equally excellent picture quality.

When the pictures are scanned at 25 or 30 per second, flicker is evident unless precautions are taken. To avoid flicker, the viewing screen is illuminated twice during each picture scan. This is nplished by interlaced scanning. *See* TELE- v SCANNING.

e aspect ratio is the ratio of the picture width s height, as scanned within the camera. This must be matched by receivers if the reprod objects are to have the correct relative hts and widths. The standard aspect ratio ughout the world is a picture 4 units wide and its high. This value was chosen to permit effit use of standard motion picture film which, r to the advent of widescreen techniques, was ed as 4 by 3 units.

lodulation standards. This second group of smission standards defines the method by which sound and picture information is carried by the ective radio carrier signals. Frequency modula- is used for the sound transmission (except in gland, France, and in some transmission in gium, where amplitude modulation is used) : sound standards define the maximum deviation of the carrier frequency and the amounts of emphasis given to the higher audio frequencies e picture information is universally transmitted

equency which is radiated by the transmitter is value determines the fineness of the picture ucture along each line (horizontally) and this ust match *approximately* the fineness of the rtical structure determined by the number of anning lines. The maximum video frequency is andardized at 3 megacycles (Mc) for the British 5-line system, 4 Mc for the American 525-line stem, 5 Mc for the 625-line European system, and 0.4 Mc for the 819 line French and Belgian systems. *See* TELEVISION.

The polarity of picture modulation must also be standardized. In negative modulation, used in the American and European systems, the power of the transmitter decreases as the scene becomes brighter. Positive modulation (power and brightness increase together) is used in England, France, and partially in Belgium. Receiver designers must adopt the polarity standard of their region; otherwise the pictures would be reproduced with tonal values reversed, as in a photographic negative. To maximize coverage for given transmitter power, the power level corresponding to full black (absence of light) in the scene is maintained constant; this standard is known as dc transmission and is used throughout the world.

with minimum interference within the transmission channel. The subcarrier sideband frequency carrying the color information is standardized relative to the picture carrier frequency to a precision of better than 1 part in 3,000,000. The number of lines per picture and pictures per second are, in turn, rigidly related to this subcarrier frequency. The color modulation standards specify the relationship between the relative intensities of the three primary colors and the reference white, and the corresponding amplitude and phase modulations of the color subcarrier. The chromaticities (color coordinates specifying hue and chroma) of the three primary colors and the reference white of the system are also standardized, as are the frequency limits associated with the subcarrier modulation components.

Channel standards. This final group of transmission standards specifies the frequency limits of the station channels and the manner in which the transmitted signals occupy the channels. The channel width must accommodate the maximum video sideband frequency as one sideband, plus a vestigial portion for the other sideband, plus room for the sound transmission and guard bands to prevent interference. Accordingly, the total channel width is 5 Mc for the 405-line British system, 6 Mc for the 525-line American system, and 7 Mc for the 625-line European system. The French system, which provides additional space for video modulation and guard bands, uses a 14-Mc channel.

Allocation standards. These standards define minimum and maximum transmitter powers in relation to antenna tower heights as well as required or permissible signal strengths at specified distances. Also specified are the minimum separation in miles and the minimum frequency separation between channels among stations which serve a given region or which are otherwise capable of causing mutual interference. In the United States, to avoid interference between receivers (the local oscillators of which can radiate interfering signals through the receiving antenna), standard values of intermediate frequency and local oscillator frequency are assumed in the allocation plan. These are closely adhered to by receiver designers. In the United States, the minimum effective radiated transmitter power is specified from 1–50 kw depending on the population to be served. Maximum powers from 100–5,000 kw are specified, depending on the channel used, the height of the antenna tower, and the geographical region in which the station is located.

Equipment standards, to assure compliance with the transmission and allocation standards, are also set up by government regulation and have been adopted, with suitable safety factors, by the trade associations. [D.G.F.]

Bibliography: D. G. Fink (ed.), *Television Engineering Handbook*, 1957.

Television studio

The basic facilities required in television studios are a lighting system, cameras, microphones, means for handling and supporting scenery, production aids such as cuing devices and titling machines, and a communications system between the control room and the studio.

TV studios range in size from small special-purpose studios, about 20 by 30 ft, used for such programs as news and weather, to large production studios, such as those used for elaborate dramatic shows. Some of these large studios are 80–100 ft wide and 150–200 ft long. Theater-type studios are used for shows with a studio audience.

High ceilings are desirable, and ceiling heights in excess of 30 ft are not unusual. High ceilings permit scenic elements to be changed quickly by hoisting one unit out of the way and lowering another. Lighting fixtures are also hung from the ceiling and may be raised or lowered.

Studio noise must be kept to a minimum. The side walls and ceilings are usually treated with sound-absorbent material to improve the acousti and to reduce the intensity of unwanted noise. A conditioning equipment must be designed and i stalled to rigid specifications for the control noise.

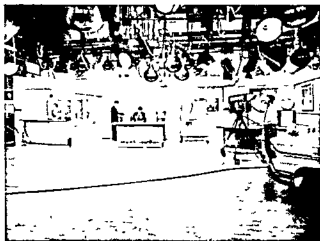


Fig. 1. General-purpose studio (*Radio Corporation of America*)

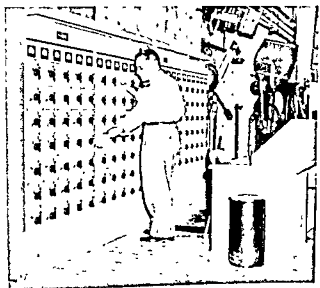


Fig. 2. Lighting control panel. (*Radio Corporation of America*)

Figure 1 shows a typical general-purpose stud. A microphone boom is used to position the micr phone for best sound pick-up and also to keep t microphone out of camera range. The two mon chrome cameras may be moved about on the sma rubber-tired wheels. Where more extensive came movement is anticipated, pedestal- or dol mounted cameras are used. *See* TELEVISION CA ERA.

Studio lighting. The somewhat limited contra range of the television system places special er phasis on television lighting (*see* TELEVISIO CAMERA TUBE). The scene will be transmitted fait fully only if the contrast range established by t lighting is within the capabilities of the televisi equipment. Small studios may have 25–50 lighti outlets each with a maximum capacity of abo 1000 watts. Large color studios often have as ma as 1000 outlets with capacities of 2–5 kw or mor In all but the smallest studios, the individual ligh ing outlets are fed from a distribution system th permits individual lights or selected groups lights to be connected to a single dimmer or switc

Figure 2 shows a lighting control panel. Son distribution systems use groups of rotary switch to connect the outlets to the required dimmer switch, others use a patch panel arrangemer where the connections are made by oversized jac and plugs, somewhat like a telephone switchboar In the small and medium-sized studios the dimme are usually manual autotransformers. Where large number of lighting changes and dimming ar switching operations are required, as in large col studios, lighting consoles are used Keys actua remote-control circuits, sometimes through ma netic amplifiers, to perform the required dimmir and switching operations. The console arrang ment brings all the lighting control functions a central point so that closer cooperation can achieved.

Control room. The control room should afford good view of the studio floor One fairly typic layout has the control room floor arranged in tw levels. The program director, the technical directo and the audio [illegible] at the higher l[illegible] the studio. They also have an [illegible] the camera monitors, which show the scenes bein picked up by all the cameras, and the line, o program, monitor which shows the picture actuall being transmitted at that instant. An intercom munications system is provided so that each di rector may contact any or all of his staff on th studio floor.

A video switching unit, located at the technica director's position, is used to switch the camera and introduce the transitions between scenes, suc as lap dissolves, superimpositions, or other elec

...tronic effects. The audio engineer operates the audio console, mixing the electrical output of the microphones to give the desired audio signal, and directs the activities of the microphone boom operator, through the intercom system, to obtain the best possible sound pickup.

The video control engineers, located at the lower level, make the necessary electronic adjustments to the camera circuits to maintain high-quality pictures.

Signal transmission. The audio and video signals from the control room may go directly to the transmitter or may go to a master control room where signals from other studios are also available. When the transmitter is located away from the studio, the audio and video signals may be sent to the transmitter by microwave circuits, or the audio signals may be sent to the transmitter by leased telephone lines and the video signals sent by leased coaxial cables or microwave circuits. [C.K.C.]

Bibliography: R. Bretz, *Techniques of Television Production*, 1953; H. Chinn, *Television Broadcasting*, 1953; R. J. Wade, *Staging TV Programs and Commercials*, 1954.

Television transmitter

...at part of a television system which converts the ...ral and visual components of a television pro...am from energy in the audio- and video-frequency ...nge to modulated radio-frequency energy capa...e of being radiated from an antenna. The term ...also applied in a broader sense to a television ...ansmitting station, including the transmitter as ...fined above, the transmitting antenna and trans...ssion line, the tower or antenna support struc...re, the aural and visual input and monitoring ...quipment, and the building in which the transmit...er and equipment are housed. The term is some...imes incorrectly applied to a television station, ...hich includes cameras and studios.

A television transmitter is required to handle both visual and aural program components simultaneously. Therefore it usually consists of two ...

... requirements and method of modulation. Signals from the two units are normally combined in a diplexer and fed into and radiated from a common antenna.

...the FCC *Rules and Regulations*.

The visual signal is an amplitude-modulated signal with the upper sideband energy extending to 4.2 megacycles (Mc) above carrier frequency and ...

...e modulation is used, so that an increase in picture brightness produces a reduction in output power. The relation between video signal levels and modulation of the visual transmitter is shown in Figure 1.

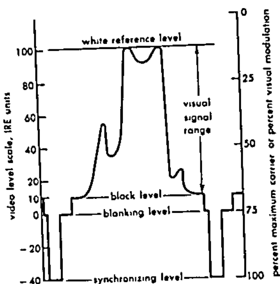


Fig. 1 Relation between video signal levels and modulation of the visual carrier.

The aural signal is a frequency modulated signal with maximum deviation of 25 kilocycles (kc). Maximum audio modulating frequency is normally not in excess of 15 kc. The aural carrier is at a frequency 4.5 Mc above that of the visual carrier. This standard spacing permits the use of intercarrier sound reception, which avoids the precise tuning that might otherwise be required in the TV receiver. See TELEVISION STANDARDS.

TV transmitters can be divided by output frequency and design characteristics into three groups, low-vhf (channels 2–6), high-vhf (channels 7–13), and uhf (channels 14–83). The transmitters in each group are designed with tubes and circuits suitable for operation at the proper frequency range.

Transmitter power. The power of signals radiated from a television station is specified in terms of effective radiated power (ERP). This is determined by multiplying the transmitter output power by the antenna power gain and the efficiency factor of the transmission line to the antenna.

Stations licensed in the United States to operate in the low-vhf channels are normally authorized an ERP of 100 kilowatts (kw). With antenna gain customary for this frequency range, visual transmitter power is usually between 10 and 35 kw. For high-vhf channels, the maximum authorized power is normally 316 kw ERP. In this case, the visual transmitter power is usually between 20 and 50 kw. For uhf stations, the authorized power may be as high as 5000 kw ERP. Transmitter power in the uhf channels is usually limited by other considerations, however, and may range from 1–60 kw.

Television transmitter power is normally specified as the rating of the visual unit. This is given

in terms of peak, or maximum, power output, which in normal transmission exists only during maximum excursion of synchronizing pulses. The aural transmitter power, which is measurable on a continuous rms basis, is normally half the visual rating. For example, a 50-kw television transmitter would have a peak visual power output of 50 kw and an aural power output of 25 kw. When transmitting a visual signal at blanking level with standard synchronizing pulses, the average visual power is slightly more than 30 kw. With normal picture transmission instead of blanking (or black level), the average power is from 10 to 15 kw On both of these conditions, the peak visual power (on sync peaks) is 50 kw.

TV antennas. Antenna power gain is determined by the size and design of the antenna and is usually higher on the ultra-high frequencies and the high-vhf channels than on the low-vhf Antenna gains range from approximately unity (gain of a dipole) to gains of 50 on the uhf band.

All television broadcast signals are horizontally polarized in the United States, meaning the signals are polarized in the same manner as those from a dipole in a horizontal position. This is considered advantageous in minimizing interference at the television receiver from automobile ignition and other impulse-type interferences, which seem to be at a maximum with vertically polarized reception. The radiation pattern of a television transmitting antenna is usually circular, in order to provide equal service in all directions from the station. Gain is achieved by concentrating the energy in the horizontal direction.

Transmitter description. A representative medium-power television transmitter as used in the United States is shown in the block diagram of Fig. 2.

The visual modulator operates from a standard monochrome or color signal This signal is amplified and used to grid-modulate the rf amplifier stage in the visual transmitter Keyed clamp-type dc insertion is employed in the modulator to establish the correct radio-frequency levels in terms of the relative video levels (*see* CLAMPING CIRCUIT). Provisions have been made for a limited amount of sync stretching and white expansion to correct for minor nonlinearity in the maximum and minimum modulation region.

The visual exciter consists of a crystal oscillator-doubler, a second doubler, a third doubler, and a tripler (*see* FREQUENCY MULTIPLIER) The output frequency is the 24th harmonic of the crystal frequency. A small variable capacitor shunted across the crystal is provided for making small adjustments of the oscillator frequency to bring it into exact agreement with the station monitor.

Three intermediate stages of power amplification are employed beyond the exciter multipliers. The second interstage power amplifier also serves as the modulated stage. This stage operates as a grounded-cathode amplifier employing grid modulation (*see* AMPLITUDE MODULATOR). Voltages on this and subsequent stages are regulated to mai[n]tain relative power levels.

The final amplifier, or output stage, employs [a] single tetrode operating in a grounded-grid or grid separation circuit, with electronically regulate[d] screen-grid voltage. The double-tuned output cir[]cuit is adjusted in normal operation to have a fla[t] response over a bandwidth of approximately 5 Mc This is achieved by careful adjustment of th[e] reactance and coupling in the output tuned cir[]cuits.

The filament of the output tube is operated fro[m] a dc source, current being supplied from a six[-]phase half-wave rectifier employing high-curren[t] germanium rectifiers.

Aural modulation is achieved by a form of pulse[-]position modulation (*see* PULSE MODULATION). A crystal oscillator at approximately 200 kc deliver[s] a saw-tooth wave, which is modulated and differ[]entiated, thus producing a pulse varying in pos[i]tion, or time, with modulation. The phase modula[]tion thus generated is multiplied many times wit[h] the multiplication of the original frequency to th[e] aural carrier frequency. Frequency modulation i[s] created as a result of correction to the audio-fre[]quency response of the phase modulator, such as t[o] be inversely proportional to frequency up to 15 kc

The aural rf amplifier circuitry is identical t[o] the visual section with the exception of the elimi[]nation of an rf stage (the modulated rf stage in the visual section). Power output of the aura[l] transmitter is 5 5 kw.

This transmitter can be connected to the antenna through the vestigial sideband filter and a diplexe[r] (combines visual and aural outputs) or it can b[e] used as a driver for a 50-kw amplifier. The vestigia[l] sideband filter attenuates the lower sideband ener[]gy to avoid interference to the next lower channel

Transmitter requirements for color. Th[e] method of color transmission developed and adopted for use in the United States is known as compatible color system This term, while applying principally to the receiver, also has meaning for other parts of the system. A television transmitter satisfying color requirements will also be satisfac[]tory for black-and-white transmission.

The system uses the same 525 lines per frame and essentially the same 30-cycle frame rate as the monochrome system. It includes chrominance infor[]mation, which is modulated on a 3.58-Mc carrier and added to the normal picture brightness or lu[]minance signals

The main points of difference between require[]ments for the transmission of color and those for black and white are in the degree of uniformity in the frequency versus amplitude response charac[]teristics, the amount of nonlinearity as a function of amplitude (measured as differential gain), and the amount of variation in phase of the chromi[]nance carrier (3.579545 Mc) as a function of am[]plitude (measured as differential phase). The fre[]quency versus amplitude characteristic must be uniform to a degree not essential in black and

hate transmission in order to maintain a standard relationship between the luminance and chromi-nance components of the signal. Phase correction is applied to the visual transmitted signal to com- [illegible] delay distortion introduced [illegible]

[illegible] re understood. It is also required that greater [illegible] ensuation be applied in the vestigial sideband filter to the lower sideband in the chrominance region 2.33 Mc below the band edge). This is to avoid interference on the adjacent channel as a result of the higher energy level at these frequencies when transmitting color. *See* COLOR TELEVISION.

Standards in other countries. The television transmitter is to a much lesser degree limited to the system for which it was designed than certain other parts of the system, such as cameras and synchronizing generators. A transmitter suitable for use on the United States 525-line, 60-field system would also serve with little adjustment or modification for the European 625-line, 50-field system. The bandwidth and modulation requirements are quite similar. It would not, however, serve without substantial modification for the English 405 line, 50-field system because the modulation in the English

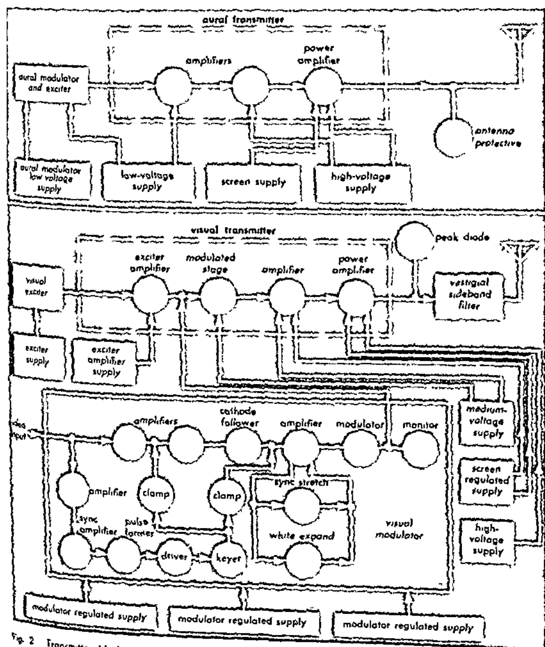


Fig. 2 Transmitter block diagram

system is positive rather than negative and the aural channel is amplitude modulated rather than frequency modulated. See TELEVISION. [R.M MO.]

Bibliography: D. G. Fink (ed.), *Television Engineering Handbook*, 1957.

Telex

A teleprinter exchange service furnished to subscribers locally and internationally by telegraph organizations in America, Europe, Africa, and the Far East.

When operated with direct dialing through automatic exchanges, users may obtain direct send and receive teleprinter connections with each other in seconds, at charges based on time and distance as recorded by automatic registers at central offices. Users also may send or receive telegrams via public message facilities at message rates

Subscribers' equipment may be arranged to prepare and transmit from perforated tape and to reperforate in tape as well as to print incoming messages.

Most Telex operation is characterized by automatic answer-back with positive machine identification codes for all subscribers This allows 24-hour service with receiving teleprinters unattended

A typical answer-back or identification device is built into each teleprinter as illustrated It is a cylindrical drum which will transmit 19 characters of teleprinter code to identify the subscriber's name and city as listed in a Telex directory. Either a trigger key marked HERE IS or a signal from a distant teleprinter key marked WHO ARE YOU will activate the drum

Having initiated a call and dialed the desired number, a subscriber of an automatic system sends the WHO ARE YOU signal to obtain automatically a printed record of the abbreviated name and city of the distant station, thus insuring that the connection is correct. He then presses the HERE IS key to identify his own station and transmits his message

Telex systems provide listing information and other central office supervisory services and include



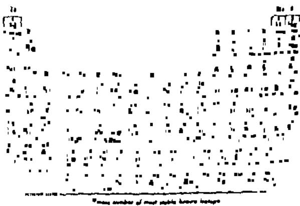
Automatic answer-back device used in teleprinter exchange service.

busy-line and stop signals, low paper safeguard, and automatic transmission testing facilities. See TELEGRAPHY; TELETYPEWRITER EXCHANGE (TWX) SERVICE. [C.H.]

Bibliography: C. J. Colombo, Telex in Canada, *Western Union Tech. Rev.*, 12(1):21-27, 1958; P. R. Easterlin, Telex in New York, *Western Union Tech. Rev.*, 13(2):45-56, 1959; General Secretariat of International Telecommunication Union, *Telex Statistics* (annually), 1955——; D. M. Rogers, The London Telex engineering control, *Post Office Electrical Engineers' Journal*, vol. 51, pt. 1, 1958.

Tellurium

Chemical element number 52, tellurium, Te, has a chemical atomic weight of 127.61. The per cent



abundances of the stable isotopes in natural tellurium are Te^{126}, 18.71%; Te^{128}, 31.79%; Te^{130}, 34.49%, Te^{120}, 0.08%; Te^{122}, 2.46%; Te^{123}, 0.87%; Te^{124}, 4.61%; and Te^{125}, 6.99%. Tellurium was first isolated by J. F. M. von Reichenstein in 1782

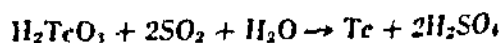
Natural occurrence. Tellurium makes up approximately 10^{-9}% of the earth's igneous rock. It is found as the free element in central Europe, Colorado, and Bolivia, and occurs with selenium in sulfur deposits in Japan It is more often found as the tellurides sylvanite (graphic tellurium), $(Ag, Au)Te_2$; nagyagite (black tellurium), $(Ag, Pb)_3(Te, S, Sb)_4$; hessite, Ag_2Te; tetradymite, Bi_2Te_3; altaite, PbTe, coloradoite, HgTe, and other silver-gold tellurides, as well as the oxide TeO_2, tellurium ocher Tellurium is also recovered from the anode slimes obtained during the electrolytic refining of copper

Preparation of the element. The extraction of tellurium is carried out by the digestion of tellurium-containing materials with hot concentrated sulfuric acid (or hydrochloric acid, for some ores) to convert the material to the tellurite (TeO_3^{--}), followed by treatment with sulfites or sulfur dioxide:

$$H_2TeO_3 + 2SO_2 + H_2O \rightarrow Te + 2H_2SO_4$$

Tellurium can also be deposited on a lead cathode from a solution of TeO_2 in hydrofluoric and sulfuric acids

Properties of the element. There are two important allotropic modifications of elemental tel-



, the crystalline and the amorphous forms. ...talline form has a silver-white color and

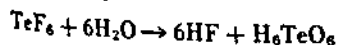
...all quantities of impurities. Light causes a small increase in electrical conductivity. ...form melts at 449.8°C and boils at 1390°C. It ... specific gravity of 6.25, a hardness of 2.5 ...ohs scale, and a Trouton constant of only ... It is insoluble in all solvents which do not ...with it, and its molecular weight at room tem...ture is not yet known. Between 1400 and ...°C, its formula is Te_2, and the Te-Te distance ...6 A. The amorphous form (brown) has a ...fic gravity of 6.015. A red colloidal sol of tel...um in water can be obtained by the reduction ...llurlc acid with hydrazine.

...ellurium burns in air with a blue flame, form... tellurium dioxide, TeO_2. It reacts with halogens ...not sulfur or selenium, and forms, among other ...ducts, both the dinegative telluride anion ...$^{--}$) which resembles selenide, and the tetraposi... tellurium cation (Te^{4+}) which resembles ...inum(IV).

...Uses. Tellurium is used primarily as an addi...e to steel to increase its ductility, as a brightener ...electroplating baths, as an additive to catalysts ... the cracking of petroleum, as a coloring ma...ial for glasses, and as an additive to lead to in...ease its strength and corrosion resistance.

Principal compounds. Hydrogen telluride (H_2Te) is the only known hydride of tellurium. A ...olorless gas with an even more offensive odor than ...ydrogen selenide, it is at least as toxic as this ...ubstance. It melts at −51.2°C, boils at −1.8°C, ...nd is less stable thermally than the hydrides of ...xygen, sulfur, and selenium, although it is a ...tronger acid in water than these hydrides. The ...iquid form has a slightly yellow color, and the ...olid is colorless. The specific gravity of the liquid at its boiling point is 2.650. The compound is decomposed by light, especially when moist, and it can be prepared by the action of acids on metallic tellurides, especially aluminum telluride, Al_2Te_3. Normal tellurides (for example, Na_2Te) are known and are less stable than the corresponding selenides toward heat. They are strong reducing agents in solution and will reduce tellurite to free tellurium. The alkali tellurides are soluble in water and are attacked by oxygen to give dark red polytellurides (for example, Na_2Te_2). Heavy metal tellurides are generally insoluble in water.

Halides. Tellurium hexafluoride, TeF_6, is a colorless gas which is formed from the elements, it melts at −37.8°C and sublimes at −38.9°C and is slowly hydrolyzed by water:



$$TeF_6 + 6H_2O \rightarrow 6HF + H_6TeO_6$$

The compound is more reactive than either its selenium or sulfur analog. Reports have been made of lower fluorides, especially tellurium tetrafluoride, TeF_4 and Te_2F_{10}. Tellurium oxyfluoride, $TeOF_2 \cdot \frac{1}{2}H_2O$, is a white crystalline material formed from anhydrous hydrogen fluoride and tellurium dioxide (other oxy halides are not well characterized).

Tellurium tetrachloride, $TeCl_4$, is formed by the action of tellurium on excess chlorine, S_2Cl_2, or $AsCl_3$; it is a white hygroscopic crystalline substance which boils at 390°C and melts at 225°C; its vapor is monomeric and orange-red up to 500°C, and it has the electrical conductivity of a salt at elevated temperatures; it is soluble in benzene, toluene, and the lower alcohols, but not in ether; its structure is a trigonal bipyramid with one equatorial position occupied by an electron pair. It reacts slowly with water to form TeO_2, and with many organic compounds, and it forms addition products with $2AlCl_3$, $2(C_2H_5)_2O$, $3NH_3$, SO_3, $6NH_3$, and $2SO_3$ per molecule of $TeCl_4$. Tellurium dichloride, $TeCl_2$, is a black solid formed by the action of tellurium on chlorine or $TeCl_4$. It melts at 175°C and boils at 324°C; the liquid has a fairly high electrical conductivity, and the solid and liquid are both reactive; water converts it to Te and H_2TeO_3, and oxygen converts it to $TeCl_4$ and TeO_2...

421°C; it is slowly hydrolyzed by excess water to TeO_2, and forms an addition compound with aniline, $TeBr_4 \cdot 2C_6H_5NH_2$. Tellurium dibromide, $TeBr_2$, is a black solid which melts at 210°C and boils at 339°C, it is hydrolyzed by water:



$$2TeBr_2 + 3H_2O \rightarrow H_2TeO_3 + 4HBr + Te$$

Tellurium tetraiodide, TeI_4, is a black solid which melts at 259°C, is slightly soluble in acetone and ethyl and amyl alcohols, and is essentially insoluble in carbon tetrachloride, carbon disulfide, ether, and acetic acid; it can be prepared by the reaction of tellurium dioxide and hydrogen iodide.

Complex halides of tellurium have also been prepared (for example, $HTeCl_5 \cdot 5H_2O$, $HTeBr_5 \cdot 5H_2O$, and $HTeI_5 \cdot 8H_2O$), the complex chloride also forms salts; in addition, salts of H_2TeX_6 have been prepared, where X = Cl, Br, and I.

Oxides. The oxides of tellurium are tellurium monoxide, TeO, tellurium dioxide, TeO_2, and tellurium trioxide, TeO_3. The monoxide is reported as a black amorphous powder which is stable in dry air in the cold, but which is oxidized in moist air to the dioxide. On being heated in vacuum, it apparently disproportionates into the dioxide and elemental tellurium. It can be formed by heating the mixed oxide $TeSO_3$:



$$TeSO_3 \rightarrow TeO + SO_2$$

and is also believed to exist at elevated temperatures in equilibrium with its decomposition products. The dioxide is the most stable oxide and is formed when tellurium is burned in air or oxygen or by oxidation of tellurium with cold nitric acid. It has two crystalline forms. One crystallizes from a nitric acid solution as colorless, tetragonal.

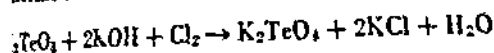
octahedronlike crystals of specific gravity about 5.8; the other, from molten tellurium dioxide as monoclinic or rhombic needles of about the same density. Tellurium dioxide melts at 452°C, and the white solid becomes a dark yellow liquid which distills at red heat, apparently without decomposition. It is only very slightly soluble in water, and forms a solution which is barely acidic, but which is soluble in concentrated acids (for example, H_2SO_4, HCl, HNO_3) in which it apparently forms salts and in strong alkalies in which it forms tellurites, such as K_2TeO_3. The dioxide is amphoteric and also forms addition compounds with strong acids, for example, $TeO_2 \cdot 3HCl$ and $(TeO_2)_2 \cdot HNO_3$. The trioxide is an orange-yellow compound formed by heating orthotelluric acid, H_6TeO_6. This oxide decomposes into the dioxide and oxygen at red heat. It is essentially insoluble in cold water, but dissolves in hot water after prolonged heating to give orthotelluric acid. It is not attacked by cold acids, but hot, concentrated hydrochloric acid converts it to a mixture of chlorine, the dioxide, and the tetrachloride. Hot, concentrated potassium hydroxide converts it to the tellurate, K_2TeO_4. There are apparently two crystal forms of the trioxide, ordinary α-TeO_3 (sp gr, 5.075) and β-TeO_3 (sp gr, 6.21) The less reactive β form is made by prolonged heating of the α form.

Acids. The important oxy acids of tellurium are tellurous acid and telluric acid. Anhydrous tellurous acid, H_2TeO_3, has never been isolated, although treatment of the potassium salt with nitric acid produces white flakes with varying quantities of water of crystallization Salts of tellurous acids from H_2TeO_3 up to $H_2Te_6O_{13}$ are known. The normal tellurites (for example, K_2TeO_3) are soluble and colorless, and tend to be oxidized to tellurates in alkaline solution by air. The acid salts (for example, $KHTeO_3$) are converted by water to the normal tellurites and tellurium dioxide. The higher tellurites are less susceptible to oxidation than the normal tellurites For example, $K_2Te_4O_9$ is not oxidized by air at 450°C. As a general rule, telluric acid resembles stannic acid. The form H_2TeO_4 has not been isolated, although its salts are known The ortho acid, H_6TeO_6, can be prepared by the oxidation of tellurium or its dioxide with chromic acid in nitric acid, with aqua regia and chloric acid, or by oxidation of the dioxide in alkaline solution with hydrogen peroxide. There is at least one additional form of the acid, allotelluric acid, or polymetatelluric acid, $(H_2TeO_4)_n$, where n is about 11. The ortho acid is the more stable form and has an octahedral configuration; it crystallizes from water as the 4-hydrate, from which the water can easily be removed In cold water it is a very weak monomeric acid which polymerizes on being heated until it becomes colloidal. This reaction is reversed on cooling. Telluric acid is more easily reduced than its sulfur and selenium analogs, and it forms normal tellurates (for example, Ag_6TeO_6), acid tellurates (also called alkaline tellurates, for example, $Na_4H_2TeO_6$, and $Na_2H_4TeO_6$), tellurate esters, such as $Te(OCH_3)_6$, and heteropoly acids and salts, such as $H_6[Te(MoO_4)_6]$. The ortho acid has two crystalline forms, a cubic form of specific gravity 3.053 and a monoclinic form of specific gravity 3.071. Nor-

Organic tellurium compounds

Type	Example	Properties
Telluromercaptans, RTeH	CH_3TeH	Bp 57°C, prep from H_2Te and RX in alcoholic NaOH
Dialkyl tellurides, R_2Te	$(CH_3)_2Te$	Bp 82°C; prep from TeX_4 and Grignards; form addition compounds, such as $(CH_3)_2Te \cdot HgBr_2$
Diaryl tellurides, R_2Te	$(C_6H_5)_2Te$	Bp 182°C at 16.5 mm; prep similar to dialkyl tellurides
Cyclic tellurides	$(CH_2)_5Te$	6-membered ring; bp 82°C at 12 mm.
	$(CH_2)_4Te$	5-membered ring; bp 166°C
Telluronium compounds, R_3TeX	$(C_2H_5)_3TeCl$	Mp 174°C; prep from dialkyl tellurides and alkyl halides; moderately strong base
	$(C_6H_5)_2CH_3TeOH$	
Ditellurides, R_2Te_2	$C_6H_5Te—TeC_6H_5$	Red crystals melting at 53°C
Aryl tellurium monohalides	RTeX	
Dialkyl tellurium dihalides	$(CH_3)_2TeI_2$	
Diaryl tellurium dihalides	$(C_6H_5)_2TeBr_2$	
Alkyl and aryl tellurium trihalides, $RTeX_3$	CH_3TeI_3	Dec. above 100°C, sol in acetone and ether to give red solutions
Telluroxides, R_2TeO	$(C_2H_5)_2TeO$	Unstable oil, forming water-soluble salts with HNO_3, formed from dialkyl tellurides and air
Tellurones, R_2TeO_2	$(CH_3)_2TeO_2$	Prep from dimethyl telluride and H_2O_2; white insoluble solid
Tellurinic acids, RTeOOH	C_6H_5TeOOH	Prep by oxid of $(C_6H_5)_2Te_2$ with HNO_3, mp 211°C
Telluric esters, $(RO)_4Te$	$(CH_3O)_4Te$	Mp 86°C prep from diazomethane and H_6TeO_6 in absolute alcohol
Telluroketones, R_2CTe	$(CH_3)_2CTe$	Bp 55–58°C at 10–13 mm, prep from H_2Te and R_2CO in HCl

al tellurates can be formed from tellurites by sing them with potassium nitrate or by passing dorine into an alkaline tellurite solution:



$$_2TeO_3 + 2KOH + Cl_2 \rightarrow K_2TeO_4 + 2KCl + H_2O$$

ellurates are reduced to tellurites by hot hydrochloric acid and to elemental tellurium by sulfur oxide. The barium salt, $BaTeO_4 \cdot 3H_2O$, is fairly oluble in water.

Mixed compounds The mixed oxide $TeSO_3$ (a ed solid) has been obtained from tellurium and ulfur trioxide It forms the telluride. K_2Te when used with potassium cyanide. Carbon sulfotelluide, S=C=Te, has been prepared by passing an rc between carbon and carbon-tellurium electrodes under carbon disulfide in the cold It is a red liquid with a melting point of −54°C and an extrapolated boiling point of about 110°C. It is quite unstable, being decomposed by light and heat.

Organic compounds. The important organic tellurium compounds are summarized in the accompanying table

See SELENIUM; SULFUR. [S.K.]

Bibliography: J W. Mellor, *A Comprehensive Treatise on Inorganic and Theoretical Chemistry*, vol. 11, 1931; H Remy, *Treatise on Inorganic Chemistry*, vol. 1, 1956, N. V. Sidgwick, *The Chemical Elements and Their Compounds*, vol. 2, 1950

Telosporidea

A class of the subphylum Sporozoa These protozoans are divided into three subclasses, the Gregarinidia, Coccidia, and Haemosporidiida. All members of the group are either intra- or extracellular parasites, and the life cycles exhibit both sexual and asexual phases. The spores lack a polar capsule and develop from an oocyst or sporocyst. The sporozoite is the usual infective stage which initiates the asexual phase in the life cycle See COCCIDIA; GREGARINIDIA, HAEMOSPORIDIIDA

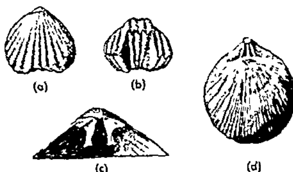
[E R BE]

Telotremata

The most specialized order of the class Articulata These brachiopods have smooth, striate, plicate or costate shells The shell substance is calcareous or fibrous prismatic, and may be punctate or impunctate The pedicle issues from between both valves

in some genera of the Spiriferacea, but may be absent or poorly developed in other genera and su-

The order is considered to have evolved from the Atremata. Five superfamilies are recognized: Rhynchonellacea (Ordovician-Recent), Atrypacea (Ordovician Devonian), Spiriferacea (Ordovician-Jurassic), Rostrospiracea (Ordovician-Jurassic), and Terebratulacea (Ordovician-Recent). This or-



(a, b) Dorsal and anterior views of Rhynchotrema capax (Ordovician, Ohio). (c) Interior of dorsal valve of *Stenochisma formosa* (Devonian, New York). (d) Dorsal view of *Trigonosemus palissa* (Cretaceous, Holland). (From W H Twenhofel and R. R. Shrock, *Principles of Invertebrate Paleontology*, McGraw-Hill, 1953)

der includes more than 200 representatives in the Paleozoic, about 165 in the Mesozoic, and 76 in the Cenozoic and Recent. Living representatives range from shallow water species such as *Hemithyris psittacea* at 3 fathoms to *Frieleia halli* which occurs in abyssal depths of 1059 fathoms. Their geologic range is from Ordovician to Recent. [K H]

Temnopleuroida

An order of Echinacea with a camarodont lantern, smooth or sculptured test, tubercles imperforate or perforate (and usually crenulate), ambulacral plates of diademoid or echinoid type, and branchial slits which are usually shallow. The order includes a long phylogenetic series in which the original characters change considerably so that a concise but exact diagnosis is not possible Following is an evolutionary summary of the three included families (1) The Glyphocyphidae, known only from the Cretaceous and Eocene and probably ancestral to the other two families, were small forms with a sculptured test, perforate crenulate tubercles, and diademoid ambulacral plates. Their sculptured test links them with (2) the Temnopleuridae whose tubercles, however, are imperforate, though usually crenulate This family arose in the Cretaceous and abounds today, especially in the tropics, on strandlines. Most of the order, so far, had shallow branchial slits, but transitional forms link them with (3) the Toxopneustidae, Tertiary and extant forms where the slits are deep and the sculpture tends to vanish. At the same time, the tubercles become imperforate, noncrenulate and the ambulacral plates change to the echinoid type. See DIADEMATOIDA, ECHINACEA, ECHINOIDA; ECHINOIDEA

[H B F.]

Temnospondyli

A *major* group of labyrinthodont amphibians in which the intercentra of the vertebral column are prominent, and the pleurocentra small to absent. Typical of the group are the Rhachitomi; also included are the Trematosauria, the Stereospondyli and, probably, the Ichthyostegalia. *See* LABYRINTHODONTIA. [A.S.R.]

Temperature

That aspect of matter which governs its ability to transfer heat from or to other matter. Temperature is related to sensory experience through two sets of nerve endings in the skin, one for warmth and the other for cold In the kinetic theory of gases, temperature is identified with the kinetic energy of translation of molecules and as such has an absolute zero where all motion is stilled. No upper limit is known, the temperature at the center of the sun is believed to be about 15,000,000°K (see Fig. 1) In classical thermodynamics, temperature is a property which, along with pressure, velocity, and so on, is used to specify the state of a thermodynamic system. The so-called zeroth law of thermodynamics holds that if two bodies are respectively at the same temperature as a reference body they will be at the same temperature as each other Heat will not transfer by conduction, convection, or radiation between two bodies at the same temperature, and will flow only from a body at higher temperature to one with lower temperature.

Empirical temperature scales. The simplest instrument for thermometry (temperature measurement) is the liquid expansion thermometer, where the length of a column of alcohol or mercury in a tube serves as an indicator Phenomena which occur at reproducible scale readings may be assigned numerical values, and readings between these values can be obtained by uniformly subdividing the scale. For example, the ice point and boiling point of pure water under constant pressure serve to define the following three scales

Scale	*Ice point*	*Boiling point*
Centigrade or Celsius (°C)	0	100
Fahrenheit (°F)	32	212
Reaumur	0	80

To convert from one scale to the other, use:

$$t_C = \frac{5}{9}(t_F - 32) = \frac{5}{4}t_{Reaumur}$$

Since the absolute lower limit of temperature can be located accurately, but never reached absolute scales have been set up with absolute zero as their base. These are the Kelvin and Rankine scales (see Fig. 2), so defined that

$$T_K = T_{C\text{-abs}} = t_C + 273.15$$
$$T_R = T_{F\text{ abs}} = t_F + 459.7$$

Many effects such as changes of electrical resistance, changes of vapor pressure, and expansion

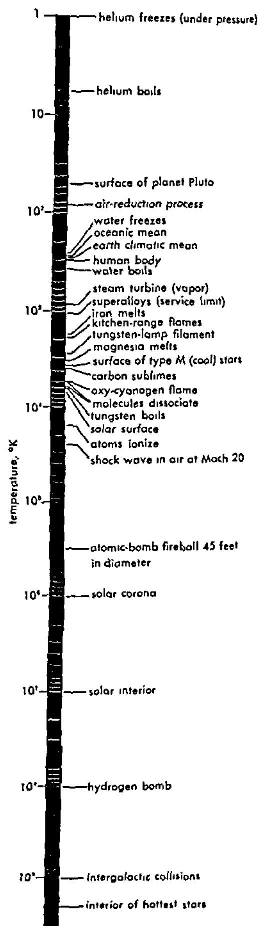


Fig. 1. Range of temperature (F. J. Dyson, What is heat? *Sci. American*, 191(3):58–59, 1954)

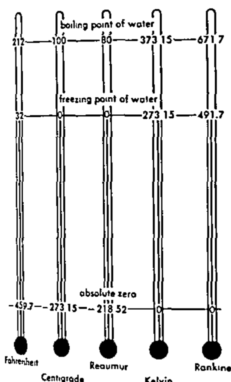


Fig. 2 Temperature scales.

of gases may serve as temperature indicators, but different thermometric substances give slightly different readings, except at the calibration points ...

each range.

Thermodynamic temperature scales. Some of the arbitrariness in the scales just described is removed when temperature is defined in terms of the second law of thermodynamics. Briefly, (1) it can be shown that *no engine* can exceed the efficiency of a reversible (Carnot) engine when both receive heat at the same temperature T_1 and reject heat at the same temperature T_2. (2) If two Carnot engines receive heat at the same T_1 but reject heat at different values of T_2, the one rejecting at the lower T_2 is the more efficient. (3) Temperature scales can be set up in *terms of* Carnot engine efficiency so long as they do not contradict statements (1) and (2). (4) One such scale gives the efficiency equation

$$\text{Carnot engine efficiency} = (T_1 - T_2)/T_1$$

This scale corresponds closely *to* the empirical absolute scales described and would be read exactly on a gas thermometer containing a so-called perfect gas. Unfortunately, neither an ideal Carnot engine nor a perfect gas exists, nevertheless, their behavior may be inferred from the behavior of real substances. Other thermodynamic scales related to this one are also valid.

The preceding statements about temperature are meaningful for systems in internal equilibrium and with many degrees of freedom. Special supplementary definitions have been, and are being, worked out for nonequilibrium states (as in explosion waves) and other special states (as when, near absolute zero, a system may have relatively few quanta of energy).

Standard temperature. By international agreement a standard temperature of 0°C has been chosen for presentation of basic physical and chemical data. For many purposes a standard nearer to room temperature is desirable; one may find 20 and 25°C, or their near counterparts 70 and 75°F, used as standards. Care should be taken to see what standard was used in a given case. *See* ABSOLUTE ZERO; BOILING POINT; CARNOT CYCLE; HEAT TRANSFER; ICE POINT; KINETIC THEORY OF MATTER; LOW-TEMPERATURE PHYSICS; LOW-TEMPERATURE THERMOMETRY; TEMPERATURE MEASUREMENT; TEMPERATURE SENSES; THERMAL EXPANSION; THERMODYNAMIC PRINCIPLES; THERMOMETER. [R.A.BU.]

Bibliography: H. C. Wolfe (ed.), *Temperature, Its Measurement and Control in Science and Industry*, vol. 2, 1955.

Temperature control, automatic

A feedback control system in which the controlled variable is the magnitude of the temperature. In many chemical, petrochemical, metallurgical, and physical processes and reactions, temperature is very critical and is carefully controlled. Temperature controllers are by far the largest single group of control devices. The temperature controller is usually set to maintain a constant temperature in the system, but there are also systems in which the temperature is made to follow some prescribed program.

In automatic temperature control systems the controlled temperature is measured by means of a temperature-measuring instrument whose output or reading is compared to a desired or reference setting. If a difference exists between the measured and desired temperature, a correction is applied to an actuator which increases or decreases the amount of heat supplied to the process. The actuator controls the flow of a heated fluid such as water, oil, a molten metal, steam, vapors, or heated gases into a heat exchanger.

An important component in temperature control systems is the element which measures the temperature at the place in the process at which control is to be applied. This device, known as a temperature transducer, converts the temperature into some other quantity, such as a mechanical movement, pressure, or electric voltage. This signal can be processed in a controller and applied to the actuator which controls the heat to the system.

Temperature transducers that produce mechanical movement are based on the difference in thermal expansion of two dissimilar metals. These are called bimetallic thermometers. Pressure-type transducers employ a closed fluid system in which a bulb is filled with a liquid or gas. A capillary

Temnospondyli

A major group of labyrinthodont amphibians in which the intercentra of the vertebral column are prominent, and the pleurocentra small to absent. Typical of the group are the Rhachitomi; also included are the Trematosauria, the Stereospondyli and, probably, the Ichthyostegalia. *See* LABYRINTHODONTIA. [A.S.R.]

Temperature

That aspect of matter which governs its ability to transfer heat from or to other matter. Temperature is related to sensory experience through two sets of nerve endings in the skin, one for warmth and the other for cold. In the kinetic theory of gases, temperature is identified with the kinetic energy of translation of molecules and as such has an absolute zero where all motion is stilled. No upper limit is known; the temperature at the center of the sun is believed to be about 15,000,000°K (see Fig. 1). In classical thermodynamics, temperature is a property which, along with pressure, velocity, and so on, is used to specify the state of a *thermodynamic system. The so-called zeroth* law of thermodynamics holds that if two bodies are respectively at the same temperature as a reference body they will be at the same temperature as each other. Heat will not transfer by conduction, convection, or radiation between two bodies at the same temperature, and will flow only from a body at higher temperature to one with lower temperature.

Empirical temperature scales. The simplest instrument for thermometry (temperature measurement) is the liquid expansion thermometer, where the length of a column of alcohol or mercury in a tube serves as an indicator. Phenomena which occur at reproducible scale readings may be assigned *numerical* values, and readings between these values can be obtained by uniformly subdividing the scale. For example, the ice point and boiling point of pure water under constant pressure serve to define the following three scales:

Scale	*Ice point*	*Boiling point*
Centigrade or Celsius (°C)	0	100
Fahrenheit (°F)	32	212
Reaumur	0	80

To convert from one scale to the other, use:

$$t_C = \frac{5}{9}(t_F - 32) = \frac{5}{4} t_{\text{Reaumur}}$$

Since the absolute lower limit of temperature can be located accurately, but never reached, absolute scales have been set up with absolute zero as their base. These are the Kelvin and Rankine scales (see Fig. 2), so defined that

$$T_K = T_{C\text{-abs}} = t_C + 273.15$$
$$T_R = T_{F\text{-abs}} = t_F + 459.7$$

Many effects such as changes of electrical resistance, changes of vapor pressure, and expansion

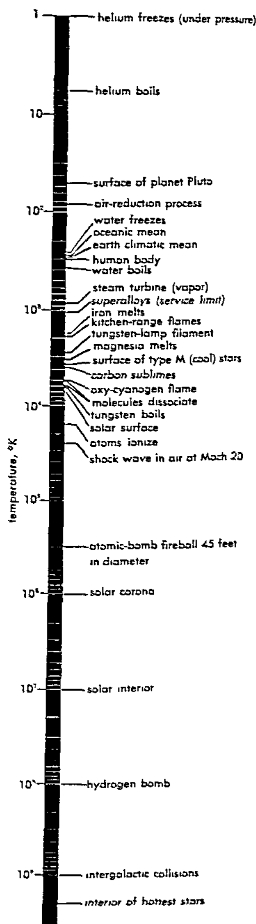


Fig. 1 Range of temperature. (F. J. Dyson, *What is heat? Sci. American*, 191(3):58–59, 1954)

7. Kinetic error: thermal element is immersed in high velocity gas.

8. Gradient error: temperature of the gas is not uniform.

The first five of the above effects are most serious in high and low-temperature measurement and when the material being measured has a low conductivity and heat capacity

Temperature measuring instruments. The instruments used for measuring temperature are known as thermometers, pyrometers, and thermocouples. The thermometer is the most common instrument at ordinary temperatures, and there are many types of thermometers, most of which can also be used as detectors in automatic control systems. By common use, a pyrometer measures temperatures above those for which thermometers are available. However, pyrometers can also be used at lower temperatures. Pyrometers are of two types. With optical pyrometers an operator visually compares the brightness of a hot object with a standard brightness for a narrow wavelength interval. Radiation pyrometers measure the rate of energy emission per unit area over a broad range of wavelengths. Radiation pyrometers can also be used as detectors in automatic control systems. The thermocouple is a unique device consisting basically of two dissimilar metallic wires joined at one end If the junction end is at a different temperature than the free ends, a voltage is developed across the junction proportional to the temperature difference between the junction and free ends The free ends can be connected to a suitable electric meter calibrated in temperature units or to a control system. See PYROMETER; THERMOCOUPLE; THERMOMETER.

Temperature indicators. In addition to these instruments, there are various other means of indicating temperatures Pyrometric cones, used mainly in the ceramics industry, soften and bend under advancing temperatures The effect is one of time and temperature and does not indicate any specific temperature (see PYROMETRIC CONE)

Temperature-sensitive crayons pellets, and paint are used as temperature indicators. Melting or color change in the material provides the temperature indication The change may be irreversible and provide a permanent indication that the calibrated temperature has been reached or exceeded Reversible units must be observed while they are undergoing the temperature cycle The paints and crayons are used to mark the object before heat is applied. Pellets are wrapped with or inserted into the package at the spot where the temperature information is desired Specific temperatures between 100 and 2000°F can be detected in this way

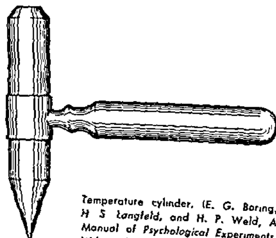
Thermography. This is a method of measuring surface temperature using luminescent materials. In contact thermography a thin layer of luminescent material is spread on the surface of an object and is excited by ultraviolet radiation in a darkened room The brightness of the coating indicates the surface temperature. In projection thermography the thermal radiation from a surface is imaged by an optical system on a thin screen of luminescent material. The pattern formed corresponds to the heat radiation of the surface. [R.E.CL.]

Bibliography: D. M. Considine (ed.), *Process Instruments and Controls Handbook*, 1957.

Temperature senses

The system of sensitivity of the skin and internal organs of the body making possible feelings of warmth and cold. It is not currently known with certainty whether there are two discrete temperature senses, one taking care of warmth, the other cold, or but a single mechanism yielding warmth under some conditions of stimulation and cold under others. The evidence for the dual-sense arrangement centers around the facts of temperature "spot" distribution and of so-called paradoxical sensations. The chief consideration favoring the single-sense interpretation has to do with the phenomena of temperature adaptation and the aftereffects of thermal stimulation.

Temperature "spots." The tip of a warmed or cooled temperature cylinder is carefully placed down in each of 100 contiguous millimeter squares of a relatively uniform and hairless region of the skin, such as the underside of the forearm. The accompanying mechanical pressure of the cylinder is kept small and constant. Some loci will respond with the appropriate sensation of warmth or cold, while others will typically yield either nothing or simply touch sensation. If arrangements are made for keeping the tip of the stimulator at a steady temperature throughout the exploration, and if precautions of timing and preservation of uniform observational set are observed, it is possible to find fair agreement between two successive mappings of the same area Agreement is rarely perfect, however, and at least a part of the instability is to be attributed to the fact that there are frequent and somewhat unpredictable shifts in skin temperature itself The blood vessels of the skin, by dilating and contracting reflexly, attempt to preserve uniformity in the face of such things as changing external



Temperature cylinder. (E. G. Boring, H S Langfeld, and H. P. Weld, A Manual of Psychological Experiments, Wiley, 1937)

tube transmits the pressure to a convenient distant point and a pressure expansive element, such as a bellows or diaphragm, converts the pressure into mechanical movement. Electrical transducers are either of the resistance or thermocouple type. In the former, the change of resistance of a metal strip is measured in a bridge circuit. A deviation from the reference value produces a proportional signal voltage. The thermocouple type consists of two wires of different metals brazed or welded together at a junction. By placing two such junctions in series, one placed at a "hot point" and the other at a "cold point," a voltage is developed which is proportional to the temperature difference between the two points. For discussion of these temperature transducers *see* THERMOCOUPLE; THERMOMETER.

The elements described above give a signal proportional to the error in temperature between the reference value, or set point, and the actual value. This error is suitably modified in a controller and applied to an actuator. Temperature controllers often are of the on-off type; they cause either full or null operation of the actuator, depending on whether the error is above or below some prescribed range of values. Actuators may be pneumatically controlled valves, which admit more or less of a heat-bearing fluid to the process, or they may be electrical devices which control the amount of current passed into a heating element.

The basic theory which describes the performance of temperature control systems is identical to that of any feedback control system Considerations of stability, steady-state accuracy, transient performance, and response to random fluctuations are important. *See* CONTROL SYSTEMS; PROCESS CONTROL [J R.R]

Temperature inversion

The increase of air temperature with height. Normally, temperature decreases with height in the lower atmosphere up to the tropopause, and then increases in the stratosphere, the "upper inversion." The rate of decrease, or lapse rate, is generally about 6C°/km or 3 3F°/1000 ft, but is somewhat less on mountain slopes and more in the free air. Inversions are caused by radiative cooling of a lower layer, by subsidence heating of an upper layer, or by the advection of warm air over cooler air or of cool air under warmer air.

Radiative exchange between the earth's surface and the sky on clear nights cools the ground and the adjacent air layer. This makes the adjacent layer colder than the layers immediately above, and creates a ground inversion a few feet to a few thousand feet thick (*see* FROST). In the most pronounced radiation inversions, over polar snowfields, and called nival inversions, temperature increases rapidly with height for 1,000–2,000 ft; above is an approximately isothermal layer several thousand feet thick; above it, normal lapse rate prevails.

Radiative cooling of the top of a cloud or dust layer creates an inversion. Sinking air warms at the adiabatic lapse rate of 10C°/km or 5.5F°/1000 ft, and can create a layer warmer than the surface layer beneath it. Cool air that displaces warm air (for instance when it blows from a cool ocean or snowfield onto warmer land) can cause a pronounced inversion that persists as long as the flow continues; likewise, warm air may flow over a cold surface layer, especially one trapped in a valley, and may cause an inversion.

Inversions effectively suppress vertical air movement, so that smokes and other atmospheric contaminants cannot rise from the surface past the inversion base. Persistent inversions are major factors in smog. *See* AIR TEMPERATURE; ATMOSPHERE; ATMOSPHERIC ADIABATIC CHANGE; ATMOSPHERIC POLLUTION; SMOG. [A C.]

Temperature measurement

The measurement of thermal potential. Two bodies are at the same temperature when there is no thermal (heat) flow from one to the other. If one body is losing heat to another, the first is at a higher temperature. Temperature scales are arbitrary but are carefully defined for precision work Temperature measuring instruments are calibrated in degrees Celsius, or Centigrade (°C), degrees Fahrenheit (°F), or degrees Kelvin (°K). The melting point of ice at standard conditions is 0°C, 32°F, and 273 15°K. The boiling point of water at standard conditions is 100°C, 212°F, and 373.15°K. *See* TEMPERATURE. For the measurement of extremely low temperatures *see* LOW-TEMPERATURE THERMOMETRY.

Temperature-measuring instruments are calibrated by comparing their readings under carefully controlled conditions with standard instruments, which in turn have been calibrated directly or indirectly against primary standards at the U.S National Bureau of Standards or a similar organization. Temperature instruments are accurate only when they are used under calibration conditions Since this is rarely possible, temperature measurement requires unusual precautions to keep errors at a minimum. Typical sources of error are the following.

1. Emergent section or transmission: a portion of the sensitive element is exposed to a different temperature for connection or reading purposes.

2. Conduction: heat is conveyed to or away from the measuring element by its supporting structure.

3. Thermal inertia: thermal elements require time to reach an equilibrium If actual temperatures are changing rapidly, thermal elements lag and do not reveal true temperature variations.

4 Radiation: in gases, thermal elements receive and transmit radiant energy from the surrounding walls Thus the temperature observed may not be that of the gas at the element.

5. Insertion: element may be improperly located, or, in small vessels, insertion of element may change the temperature of the body.

6. Conversion: temperatures can be observed by conversion to mechanical or electrical signals.

7. Kinetic error: thermal element is immersed in high velocity gas.

8. Gradient error: temperature of the gas is not uniform.

The first five of the above effects are most serious in high- and low-temperature measurement and when the material being measured has a low conductivity and heat capacity.

Temperature measuring instruments. The instruments used for measuring temperature are known as thermometers, pyrometers, and thermocouples. The thermometer is the most common instrument at ordinary temperatures, and there are many types of thermometers, most of which can also be used as detectors in automatic control systems. By common use, a pyrometer measures temperatures above those for which thermometers are available. However, pyrometers can also be used at lower temperatures Pyrometers are of two types. With optical pyrometers an operator visually compares the brightness of a hot object with a standard brightness for a narrow wavelength interval. Radiation pyrometers measure the rate of energy emission per unit area over a broad range of wavelengths Radiation pyrometers can also be used as detectors in automatic control systems. The thermocouple is a unique device consisting basically of two dissimilar metallic wires joined at one end If the junction end is at a different temperature than the free ends, a voltage is developed across the junction proportional to the temperature difference between the junction and free ends. The free ends can be connected to a suitable electric meter calibrated in temperature units or to a control system. *See* PYROMETER; THERMOCOUPLE; THERMOMETER.

Temperature indicators. In addition to these instruments, there are various other means of indicating temperatures. Pyrometric cones, used mainly in the ceramics industry, soften and bend under advancing temperatures. The effect is one of time and temperature and does not indicate any specific temperature (*see* PYROMETRIC CONE).

Temperature sensitive crayons, pellets, and paint are used as temperature indicators Melting or color change in the material provides the temperature indication. The change may be irreversible and provide a permanent indication that the calibrated temperature has been reached or exceeded. Reversible units must be observed while they are undergoing the temperature cycle The paints and crayons are used to mark the object before heat is applied. Pellets are wrapped with or inserted into the package at the spot where the temperature information is desired Specific temperatures between 100 and 2000°F can be detected in this way

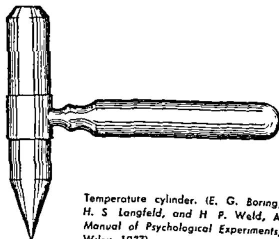
Thermography. This is a method of measuring surface temperature using luminescent materials In contact thermography a thin layer of luminescent material is spread on the surface of an object and is excited by ultraviolet radiation in a darkened room. The brightness of the coating indicates the surface temperature In projection thermography the thermal radiation from a surface is imaged by an optical system on a thin screen of luminescent material. The pattern formed corresponds to the heat radiation of the surface. [R.E.CL.]

Bibliography: D. M. Considine (ed.), *Process Instruments and Controls Handbook*, 1957.

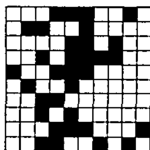
Temperature senses

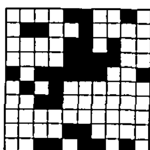
The system of sensitivity of the skin and internal organs of the body making possible feelings of warmth and cold. It is not currently known with certainty whether there are two discrete temperature senses, one taking care of warmth, the other cold, or but a single mechanism yielding warmth under some conditions of stimulation and cold under others. The evidence for the dual-sense arrangement centers around the facts of temperature "spot" distribution and of so-called paradoxical sensations. The chief consideration favoring the single-sense interpretation has to do with the phenomena of temperature adaptation and the aftereffects of thermal stimulation.

Temperature "spots." The tip of a warmed or cooled temperature cylinder is carefully placed down in each of 100 contiguous millimeter squares of a relatively uniform and hairless region of the skin, such as the underside of the forearm. The accompanying mechanical pressure of the cylinder is kept small and constant Some loci will respond with the appropriate sensation of warmth or cold, while others will typically yield either nothing or simply touch sensation. If arrangements are made for keeping the tip of the stimulator at a steady temperature throughout the exploration, and if precautions of timing and preservation of uniform observational set are observed, it is possible to find fair agreement between two successive mappings of the same area Agreement is rarely perfect, however, and at least a part of the instability is to be attributed to the fact that there are frequent and somewhat unpredictable shifts in skin temperature itself The blood vessels of the skin, by dilating and contracting reflexly, attempt to preserve uniformity in the face of such things as changing external



Temperature cylinder. (E. G. Boring, H. S Langfeld, and H P. Weld, A Manual of Psychological Experiments, Wiley, 1937)





Cold spot patterns. Successive mappings, *a* and *b*, 5 min apart, of a 1-cm² area on the forearm. Blackened squares indicate reports of cold. *(E. G. Boring, H. S. Langfeld, and H. P. Weld, Foundations of Psychology, Wiley, New York, 1948)*

thermal conditions and evaporation of skin moisture Since the amount of heat carried away from or delivered to the skin by a temperature stimulator is in no case great, and since the epidermis is a good heat insulator, it is not surprising if a particular spot sometimes responds, and at other times does not. The apparent fickleness of temperature spots may thus be the simple consequence of high absolute sensitivity of the temperature-reporting mechanism coupled with a fair degree of autonomy of temperature regulation by the skin itself. Even though such instability is encountered when whole areas are mapped, it should be said that occasionally spots are found which respond with the proper sensation whenever stimulated, even though testing and retesting may be weeks or even years apart.

However poor reliability of measurement may be in the mapping of temperature spots, it is still the case that there are quite characteristic pattern differences between plots of cold spots and those of warm spots. In fact, the patterns are so different as to bring into question the term spot, where warmth is concerned. There are small warmth areas, larger than cold spots and therefore less dense in distribution. A comparison of warm- and cold-spot concentration always shows that the cold greatly outnumbers the warm. Moreover, the patterns are never close to being coincident, and it is this fact that presents strong evidence for the interpretation that there are two different sets of nerve endings for the two systems of sensitivity, hence two temperature senses.

Paradoxical cold. Whereas withdrawal of heat from the skin is the normal or adequate stimulus for cold sensations, it is also possible to evoke them by other means. Low-frequency alternating currents, applied to the skin through relatively small electrodes, sometimes arouse cold vibration. Simple mechanical, nonthermal stimulation, such as the movement of a hair projecting from the skin, may elicit a flash of cold. Still another nonadequate stimulus for cold is heat itself. It occasionally happens, in exploring with a warm stimulus, that a spot normally responsive to cold will react with a cold sensation. Since it is paradoxical that a hot stimulus should give a perception of cold, this phenomenon has been called paradoxical cold The range of temperatures over which this effect may be produced is roughly from 42–50°C (108–122°F). At about 52°C (126°F) the stimulus is hot enough to produce injury to the tissue and evoke *thermal pain*. The importance of paradoxical cold lies in the support it gives to the dual theory of the temperature senses. It also supports the doctrine of specificity of nervous energy. *See* SENSATION.

Thermal adaptation. All kinds of sensations, including those of temperature, display the phenomenon of adaptation in accordance with which steady stimulation results in a gradual decline of intensity and, in many instances, eventual disappearance of the sensation. A hand plunged into cold water (if the temperature is not too low) feels at first cold, then less so, and ultimately the cold will fade away completely. The same happens with warmth. Everyone is familiar with the fact that one gets used to a too-cold shower or a too-hot tub. Actually, there are limits within which total loss of thermal sensation will occur. Cold stimuli below 16°C (61°F) and hot ones above 42°C (108°F) will not fade out completely in extended cutaneous areas, such as a hand or arm. Cutaneous spots behave differently.

The temperature to which the skin is adapted, at any given moment, is called physiological zero. This thermal condition gives rise to neither warmth nor cold. Let an appreciably warmer stimulus now be applied and, through adaptation, physiological zero shifts upward. Conversely, cold moves the zero downward. A 33°C (91°F) temperature, which normally yields thermal indifference, feels cold following 40°C (104°F) adaptation and warm after 25°C (77°F) stimulation. Important for its bearing on the theory of the mechanism involved is the fact that both kinds of sensitivity move together. It is argued that if cold adaptation affects thresholds for warmth, and vice versa, a single temperature sense must be responsible for the two

Temperature receptors. There is currently no sure knowledge as to the nervous mechanism responsible for sensations of cold and warmth. The classical theory, that of Max von Frey, is that cold is mediated by Krause's end bulbs and warmth by Ruffini cylinders (*see* CUTANEOUS SENSATION). This presumed correlation is based on (1) the shorter reaction time to cold and more superficial location of the Krause ending, and (2) good agreement between sensitivity and end-bulb distribution in the sclera of the eye. Attempts to chart sensitivity, then to excise the tissue and examine it histologically, have been largely disappointing, since there is almost uniform failure to find the expected specialized nerve terminations

A [illegible]

J. P. [illegible]

tion. [illegible]

perature sensitivity, both the cold and warmth patterns are handled by nerve endings in the smooth muscle walls of the tiny arterioles of the skin. Cold is a consequence of contraction of blood vessels.

warmth the result of their dilation. Thermal pain, occurring both at high (52°C or 126°F) and low (3 C or 37°F) temperatures, comes from spastic contraction of the same muscle elements. Like the classical theory, the neurovascular one explains [illegible]

ever, since they are far apart in basic conception.

Studies of the electrical changes occurring in nerves of the cat's tongue when the tissues housing the endings are warmed or cooled point to a clear separation of the two kinds of sensitivity. "Warm" and "cold" fibers act differently and are of different size and conductive properties. Moreover, impulses may be generated by direct cooling or warming of the nerve axon, thus suggesting that specialized endings of the Krause or Ruffini type may not be a vital part of the picture. [F.A.C.]

Bibliography: M. A. Wenger, F. N. Jones, and M. H. Jones, *Physiological Psychology*, 1956

Temperature-humidity index

Formerly called a discomfort index, the temperature-humidity index was developed by the U S Weather Bureau to give a single numerical value, in the general range of 70–80, which would reflect the outdoor atmospheric conditions of temperature and humidity as a measure of comfort (or discomfort) during the warm season of the year. Temperature-humidity index I_{TH} is defined as

$$I_{TH} = 0.4\,(\text{dry-bulb temp, °F} + \text{wet bulb temp, °F}) + 15$$

When the index is 70 practically all people feel comfortable; when it is 80 no one is comfortable; and when it is 75 about half the population is satisfied. The index is useful in the prediction and allocation of power-system loads resulting from the impact of air-conditioning equipment operation. See DEGREE DAY; PSYCHROMETRICS [T B]

Tempering

A process whereby hardened and quenched steel, in the form of tools, springs, and certain types of wire, is reheated and then cooled again at various rates to decrease the brittleness and increase the toughness of the metal. In tools, often only the cutting edge and the part adjacent to it are tempered. The degree of tempering is determined by the color the steel turns as it is heated, from pale yellow (roughly 430°F) for steel engraving tools and hammer faces, to dark purple (approximately 550 F) for saws, springs, needles and the like When the part being tempered reaches the proper color for the toughness required, it is again quenched or otherwise cooled. Tempering is also sometimes called drawing See HEAT TREATMENT (METALS AND ALLOYS). [G.CO]

Tendon

A white, glistening, fibrous cord which joins a muscle to some movable structure such as a bone or cartilage. A tendon permits concentration of muscle force on a small area, allows muscles to act at a distance, and, in some cases, changes the direction of pull, thereby increasing leverage. Flattened, sheetlike tendons found on broad muscles are called aponeuroses. Tendons consist almost entirely of white fibrous connective tissue formed by closely packed, parallel bundles of collagen fibrils. Although almost nonelastic, tendons are remarkable for their pliability and toughness. Considerable variations exist in shape and size, which depend on the location, function, and power of the associated muscle.

Surrounding the tendon is the tendon sheath, or vagina fibrosa, usually a dense, fibrous, connective-tissue layer. Between the tendon bundles and the sheath is a fine network of small lymphatic vessels. This tendon space also contains a small amount of fluid, similar to that in joint cavities, but occasionally this space is filled with a more viscid, mucous material. The fluid and the sheath appear to facilitate the movement of the tendon, reducing friction and irritation. The sheath is normally the means by which the tendon is held in place; small connective-tissue projections intermingle with surrounding tissue, thus binding the entire structure in place, yet permitting a freedom of movement of the tendon proper See MUSCULAR SYSTEM, SKELETAL SYSTEM [E G ST.]

Tensor

A set of components that are functions of a point in every coordinate system under consideration. Tensors can be transformed linearly and homogeneously from one coordinate system to another. *See* CALCULUS OF TENSORS

Tentaculata

A class of the phylum Ctenophora consisting of the four orders: Cydippidea, Lobata, Cestida, and Platyctenea These animals are characterized by having tentacles which may be variously modified not only structurally but functionally as well. *See* CTENOPHORA, *see also* CESTIDA, CYDIPPIDEA; LOBATA, PLATYCTENEA.

Teratogenesis

The development of a fetal monstrosity. Abnormal individuals human and animal, have been known since antiquity and have been variously regarded and interpreted as visitations of the displeasure of gods, or as a result of unnatural matings or other unfounded ideas. It is now recognized that teratogenesis results from some error of development. The formation of a normal individual depends upon the proper synchronization of a well-ordered sequence of chemical and morphological events. During the course of development there are many opportunities for dislocation of the synchrony of these processes followed by more or less dire consequences.

Abnormality. It is easy to recognize the markedly abnormal individual No one questions the classifi-

cation of a two-headed individual as a teras. The anatomy of very few individuals conforms to the "normal" pattern because there are always slight variations in size of structures and in their disposition and location. As long as the discrepancies are not marked, and function is not impaired, the individual is considered normal. The demarcation between the almost normal and the definitely abnormal is not always sharp and frequently the decision as to whether a border-line case is to be classed abnormal is an arbitrary one.

Scientific interest. Scientific interest in teratogenesis has continued throughout the period of recorded biology and medicine and has become intensified since World War II for several reasons. (1) The likelihood of increased radiation exposure for the population at large has made it imperative that more be known about the direct effects of

[illegible]

of congenital anomaly such as cataract, deafness, and mental deficiency among the children of women who have been infected with rubella, or German measles, during the first 2–3 months of pregnancy. This has opened the possibility that other diseases may have similar consequences (3) Certain immunological incompatibilities between the mother and fetus may produce anomalies in the fetus (for example, erythroblastosis fetalis caused by Rh incompatibility). (4) Infant mortality statistics indicate that there is a relative increase in the incidence of congenital malformations among children from ages 1–4 years. This is the result of three factors: the decline of deaths due to infectious disease; improved recognition of certain types of defects as congenital; and, paradoxically, improved prenatal care and nutrition. Some malformed children which formerly would have died in utero now come to term (5) Comparison of the details of anomalous development with the normal provides the embryologist with a tool for elucidating normal developmental mechanics.

Causation. There are several levels at which answers to the cause of malformations may be sought One concerns the kinds of agents involved Another is concerned with the mechanics of action of the agents. Again, the latter may be studied at two levels: one concerned with the visible developmental mechanics and the other with the less well understood biochemical and biophysical processes on which the more evident phenomena are dependent. In other words, cause may mean different things depending on point of view.

Causal agents. Factors responsible for the development of anomalies may be separated into genetic and nongenetic or environmental. However, this separation is a utilitarian one because there is less distinction between the two groups than seems evident. Whether a given individual will succumb to a particular environmental insult may depend on its genetic make-up. There is a well-established genetic basis for susceptibility to a variety of agents, including those responsible for anomalie

Genetic factors. Geneticists have studied man conditions in laboratory animals which are cause by the action of mutant genes. The effects may b noted early or late and may be profound, that i lethal, or trivial, depending upon the system altere and the degree to which it is affected. Both dom nant and recessive factors may be involved. Man mutant conditions in *Drosophila*, as well as i other laboratory and domestic animals, are terat logical. There are also a number of human anom lies, such as polydactyly, syndactyly, achondr plasia, cleft palate, microphthalmia, spina bifid which have a known genetic basis. It is suspecte that others also have a major genetic involvemen

Nongenetic factors. Anomalies have been pr duced in laboratory animals by a vast array chemical and physical agents. Heavy metals, alc hol, excess or deficiency of hormones, oxygen lac nutritional conditions including vitamin excess deficiency, metabolic inhibitors, virtually a chemical, and irradiations of various types a among these. Virus infections may act directly on fetus or indirectly through alteration of matern metabolites. The array of agents is so vast that may be safe to say that virtually any substance physical agent may produce anomalies when a ministered in a particular way to a particul embryo. The nature of the anomaly may also va greatly, and it depends on the agent used, stag of development of the organism when it is fir exposed to the agent, the dose and duration of th insult, and the genetic make-up of the exposed i dividual The physiological condition of the i dividual, either germ cell or embryo, at the time exposure may be important.

Essentially the same condition may be produce by several independent mutations as well as by number of experimental treatments. Details development which lead to a given condition ma differ, but the end results are indistinguishabl As an example, rumplessness in chickens (absenc of the tail) may be the result of the action of tw well-known mutations, one dominant, the other re cessive, as well as of a variety of experiment insults It is difficult to recognize the agent fro the appearance of any one individual. Know edge of pedigree and of the entire developmenta history of the individual is required for this. Th last statement must be emphasized, because th involves one of the most troublesome problems i medicine. Physicians must frequently decid whether a given anomaly, occurring in a single in dividual, is inherited, and estimate the likelihoo that it will appear in subsequent children. Answer must be based on knowledge that the particula condition has a proven genetic basis in othe families and on a complete study of the family i question. When such information is lacking ther is recourse to published data which indicate th statistical probability of a second occurrence.

Age at treatment. The stage of an embryo at th time it is insulted is important in determining th effect of the insult. In ge nen structur

most susceptible when it is in its earliest forma-
te condition and deleterious treatment at this
age will have its most drastic effect. This gener-
lization does not always hold. Sometimes people
apply it in trying to establish when a given condi-
on may have been initiated. *This may be mis-*
leading. Limblessness may result from a failure of
limb to form. The earliest limb bud may be pres-
nt but does not develop into a complete append-
age. In other cases a complete limb may form and
subsequently degenerate. The stage showing the
initial manifestation is quite different in the two
cases, but the end results may be similar.

nature of the treatment was considered unimpor-
tant. However, some agents such as chemical
substances have the same effects even though ad-
ministered over a wide range of stages. These
substances are selective in their effects. Treatment
of a younger embryo may result in a more drastic
anomaly, *but of the same type encountered after
treatment* of an older one. This does not imply
that the substance has a specific reaction with only
the responding tissue. It is more likely that all tis-
sues react, but that the affected structure is most
sensitive to the agent. Its development is altered
whereas that of other structures is not.

Production of effects. Studies of genetic and
experimental anomalies have revealed that there
are no basic differences between them. That is,
there is only one set of developmental principles
and whether the insult which results in a distorted
developmental sequence is gene-mediated or the
result of environmental alteration, the conse-

development of the anomaly. This is impractical
for humans and many large animals (laboratory
animals, especially *mice and chickens*) are used
for such studies. Some generalizations and some
examples follow. Anomalies *result from some type*
of distortion of a normal developmental process,
and all developmental processes are subject to
disturbances of some sort. Developmental phe-
nomena and their disturbances may be broadly
grouped as follows.

Form changes. This may include interference
with tissue movements. The *earliest morphogenetic*
movements involve shifts of tissues which result in
the establishment of the body axis. All subsequent
normal development depends on the normality of
these events. *Such* movements may either be
blocked completely, so that no embryo develops,
or they may be distorted so that the embryo is
deranged. More localized disturbances of tissue
movements may result in less drastic anomalies.
For example, pigment patterns may be atypical
if propigment cells do not migrate to the proper
place. See ANIMAL MORPHOGENESIS.

Tissue interactions. This includes a wide array
of possible disturbances. The inductive stimulus
may be deficient or excessive; responding tissue
may give an incomplete or exaggerated response.
There may be mechanical interferences between
the two tissues, that is, when an abnormal amount
of fluid accumulates between them. In some in-
stances both stimulus and response are defective.
See EMBRYONIC INDUCTION.

Differentiation. The differentiation of a given
tissue depends on many preceding events. Forma-
tion of brain depends on the transmission of the
proper stimulus from the underlying tissue to the
prebrain tissue. Other requirements may be dis-
turbed; some tissues require a given concentra-
tion of cells and will not form when fewer com-
petent cells are present. As an example, cartilage
does not form in some cases of limb anomaly, in-
stead a ligament forms in place of the long bones
of the limb at the site of cell deficiency. *See* EM-
BRYONIC DIFFERENTIATION.

Growth processes. By growth is meant increase
in size. There may be interference with growth of
the entire organism, as in dwarfism or gigantism,
or of some structures. In some types of dwarfism
all structures are reduced proportionately. In
others the body may be normal but the limbs small,
that is, disproportionate dwarfing. Individual struc-
tures may be reduced, such as the eyes, digits or
limbs. Some types of dwarfism and gigantism have
been traced to malfunctioning pituitary glands. In
others the cause is obscure. See PITUITARY GLAND.

Degenerative changes. Certain types of degenera-
tion are a normal part of development; in other
cases there are degenerative changes where none
should occur. Anomalies may arise from a failure
of normal degeneration to occur in instances such
as webbing between toes or persistent heart ducts
in some types of blue babies. They may arise when
normal degenerative processes become more ex-
tensive than usual. For example, in one type of
genetic rumplessness of chickens the tissue which
should form the tail degenerates because of the
spreading of a nearby center of degeneration which
normally is involved in development of the anus.

Sequential anomalies. Many mutant genes or ex-
perimental conditions produce more than one ef-
fect. In many instances these disturbances may be
traced back to a single effect which leads, in a
series of steps, to the many eventual effects. Some
geneticists draw up a pedigree of causes in which
the multiple effects are related to the original.
There are, however cases *in which multiple* effects
are not traceable to single causes by ordinary pro-
cedures. These may be the result of factors which
act on more than one primary process whose rela-
tion to the several morphological consequences is
not completely known.

Premorphological events. It is assumed that all
morphological events are the visible consequences
of prior biochemical and biophysical events. From
this point of view it is *cogent to inquire* into these
aspects of the various phenomena discussed above.
What are the biochemical disturbances which alter
activity or reactivity of tissues, which produce
degenerative changes? It is in this area that knowl-

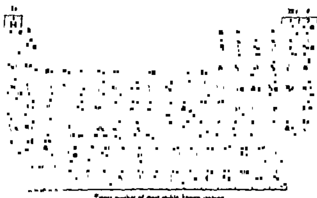
edge is most weak; a reflection of lack, at this level, of knowledge about normal events. The use of consistently produced anomalies, by genetic or experimental means, may prove a valuable tool for obtaining this type of information. Genetic factors are now known to alter eye color in *Drosophila* by blocking of particular steps in a synthetic sequence. Many teratogenic agents are known to have effects on particular pathways in carbohydrate metabolism. A key metabolite may completely mitigate the effects of a teratogenic agent. A number of antimetabolites have been shown to produce effects which are eradicated by the proper dose of the metabolite The example of sickle-cell anemia is suggestive of the type of information which must be sought. In this inherited condition of humans, red blood cells have an abnormal or sickle shape and do not function normally. The atypical cell shape is related to comparatively minor changes in the structure of the hemoglobin molecule which are the consequence of the change of one amino acid residue out of about 300. One particular glutamic acid residue is replaced by valine and the effects on structure and function of the cells are profound. *See* EMBRYOLOGY, EXPERIMENTAL.

[E.Z]

Bibliography. J Warkany, Conference on teratology, *Pediatrics*, 19:719–792, 1957; B H Willier, P. A. Weiss, and V Hamburger, *Analysis of Development*, 1955.

Terbium

Element number 65, terbium, Tb, is a very rare metallic element of the rare-earth group. Its atomic

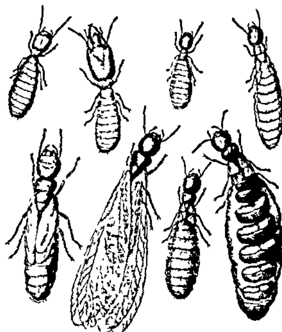


weight is 158.93, and the stable isotope Tb^{159} makes up 100% of the naturally occurring element. It was discovered in 1843 by C G. Mosander, who originally named the oxide terbia, but it has been known as terbium since 1877. The element was first isolated in fairly pure form by G. Urbain in 1905. The common oxide, Tb_4O_7, is obtained when its salts are ignited in air. Its salts are all trivalent, white in color, and when dissolved give colorless solutions. The only quadrivalent form of terbium known is in the higher oxides, and if Tb_4O_7 is ignited under a high pressure of oxygen, a compound approaching the composition TbO_2 can be prepared. The higher oxides slowly decompose when treated with dilute acid to give the trivalent ions in solution. For properties of the metal, *see* RARE EARTH ELEMENTS. Although the metal is attacked readily at high temperatures by air, the attack is extremely slow at room temperatures. The metal is paramagnetic and has a Curie point in the neighborhood of 230°K.

[F.H.SP.]

Termite

Any member of the insect order Isoptera. There are about 1800 species, most of them tropical. Several



The termite, Reticulitermes flavipes (From P. Martin Duncan, ed., Cassell's Natural History, Cassell)

species occur in the United States. Some tropical species build elaborate nests above ground, but North American forms live underground.

Termites, frequently called white ants, superficially resemble ants but can be recognized in any stage by the broad connection between the thorax and abdomen, this junction is constricted sharply in the ants Workers and soldiers are wingless. Winged, sexual forms swarm in the spring. They have two pairs of similar membranous wings which are carried flat on the back when at rest. After swarming, sexuals settle in pairs, bite off their wings, and start a nest. Each female, or queen, lives several years and produces millions of eggs Their metamorphosis is gradual.

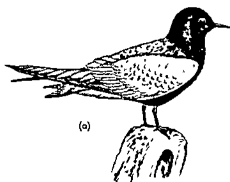
Workers may range far from the nest in search of food, always traveling in subterranean tunnels. They eat wood and frequently cause great damage to houses, furniture, utility poles, fence posts, and other wooden structures Most termites have symbiotic colonies of Protozoa in their intestines which aid in the digestion of cellulose *See* ISOPTERA.

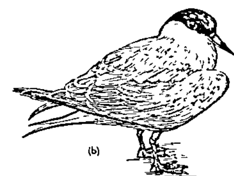
[J D B]

Tern

Any of about 50 species of moderate-sized, fish-eating birds belonging to the subfamily Sterninae, family Laridae, a family which also includes the

...lls and skimmers. Terns are similar to gulls, but ...y be distinguished in flight because terns carry ...e bill turned downward rather than straight ...ead Terns are littoral birds, staying near the ...ore at all times. They are usually smaller and ...ore slender than gulls and do not alight on the ...ater to rest. Most of the terns are marine, but ...veral species may also be found nesting near ...rge bodies of fresh water; one, the black tern, *...hlidonias nigra*, is entirely a fresh-water species.





Terns. (a) Black tern, *Chlidonias nigra*, length to $10\frac{1}{4}$ in. (b) Common tern, *Sterna hirunda*, length to 16 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

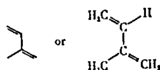
Many terns have long tails, sometimes very long and deeply forked Terns are commonly white or gray; several have black crowns. There are 12 species of terns in the United States Most of them are widely distributed, a few are virtually cosmopolitan. See CHARADRIIFORMES; GULL [J D B]

Ternifine man

A fossil human type known by three jaws and a parietal bone Discovered in 1954 and 1955 at the village of Palikao, 10 miles southeast of Oran Algeria, all the parts derive from sands of a former lake floor They were associated with a fossil fauna of the beginning of the middle Pleistocene and with early Acheulian stone tools, suggesting a date of approximately Second Glacial age. The jaws strongly resemble those of Pekin man from Choukoutien in size and in details of form (thickness of symphysis, receding chin, form of ascending ramus). The teeth show similar resemblances, with robust premolars having traces of basal cingulum C. Arambourg, the discoverer, named the type *Atlanthropus mauritanicus*. See FOSSIL MAN; PEKING MAN. [W.W H]

Terpene

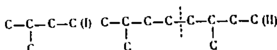
One of a class of chemical compounds having a structural relationship to isoprene.

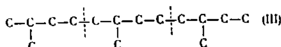


Isoprene

The name terpene is applied not only to hydrocarbons but also to their oxygenated, and other, derivatives. Most terpenes are unsaturated acyclic and cyclic compounds characterized by chemical instability. The aliphatic terpenes readily form ring structures Most terpenes are easily oxidized, reduced, and polymerized. They undergo intramolecular rearrangements and shifting of the olefinic bonds They possess distinctive odors; the majority show optical activity, and they are usually colorless liquids difficult to isolate by fractional distillation

Terpene classification. The terpenes are classified on the basis of multiples of the 5-carbon isoprene structure The skeleton (I) is referred to as a hemiterpene, (II) as a monoterpene, (III) as a

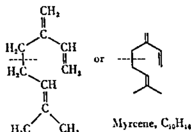




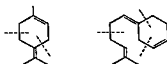
sesquiterpene Diterpenes contain four isoprene units, triterpenes six, tetraterpenes eight, and polyterpenes a large number, *n* units.

The fusion of these isoprene units may produce open branched-chain, or cyclic compounds such as:

A terpene having the empirical formula $C_{10}H_{16}$ indicates unsaturation either because of the pres-



Myrcene, $C_{10}H_{16}$



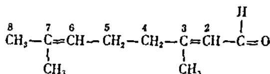
Terpinolene, $C_{10}H_{16}$ γ-Bisabolene, $C_{15}H_{24}$

ence of olefinic bonds or cyclic structure. Myrcene is acyclic and has three olefinic bonds; terpinolene is monocyclic and contains only two olefinic bonds.

Terpenes are further classified according to the number of rings in their structure. Some are open chain, many are monocyclic, and others contain three and four rings. Not all rings are six-membered.

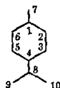
The oxygenated forms of the terpene hydrocarbons are an important group of perfume and flavor products.

Except for their trivial names, acyclic terpenes are most conveniently named by the rules of systematic nomenclature. For example, citral would be designated 3,7-dimethyl-2,6-octadiene-1-al. The

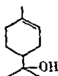


Citral

monocylic terpenes are conveniently named on the basis of the structure called menthane and numbered as follows:



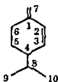
Menthane



α-Terpineol

α-Terpineol would then be named *p*-menthene-1-8-ol.

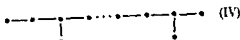
Beta-phellandrene would be named *p*-menthadiene-1(7)2. It is often necessary to indicate the direction of the olefinic bond by bracketing the proper carbon atom such as 7, 8, 9, or 10



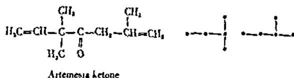
β-Phellandrene

If the olefinic bonds of *p*-menthadiene are rearranged, a total of 32 isomers will be formed, including optical isomers, racemic forms, and cis-trans forms. A number of these forms exist in nature.

Most of the terpenes are constructed on a *head-to-tail combination* of isoprene units (IV). Occa-

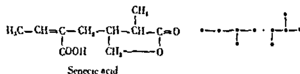


sionally this is not *true; for example,*

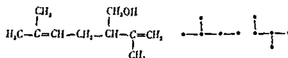


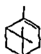
Artemesia ketone

The union of isoprene units is further varied by ring systems present in naturally occurring terpenes and those formed by internal rearrangement.

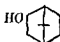


Senecic acid

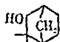




α-Pinene



Borneol



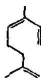
Fenchyl alcohol



Carene-1

Monoterpenes. These terpenes include acyclic, monocyclic, dicyclic, and tricyclic compounds.

Acyclic monoterpenes. The hydrocarbons of this class are not important commercially, nor are they

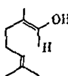


Ocimene

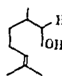


Myrcene

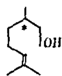
numerous Myrcene and ocimene are examples. The alcohols, geraniol, citronellol, linalool, and nerol, and their esters are widely used in perfumes. They occur in many essential oils. The asterisks indicate asymmetric carbon atoms, the presence of which makes the compounds optically active (*see* OPTICAL ACTIVITY)



Geraniol



Citronellol



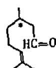
Linalool



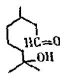
Nerol

The aldehydes citral and citronellal occur in nature. Citral is a starting material for the manufacture of the ionones (terpenes containing the ketone group) and is also used for its lemon flavor. Citronellal can be reduced to citronellol and can also

ated to yield an important perfume mate-
roxycitronellal



tral



Citronellal

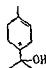
Hydroxycitronellal

nocyclic monoterpenes. The most important ercial hydrocarbon of this class is limonene.

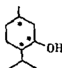


Limonene

e chief constituent in orange, grapefruit, and non oils is *d*-limonene. Orange oil is over 90% imonene. *l*-Limonene occurs in smaller amounts

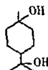


Terpineol



Menthol

eol occurs in the optically active and racemic as the alcohol and as esters. It has been d in the oils of petitgrain, neroli, and camphor in pine oils The latter is a fraction of the e oils obtained from pine stumps The perfume



Terpin hydrate

ade is obtained commercially from terpin hydrate hich is formed by the hydration of pinene.

Dicyclic monoterpenes. For discussions of the ost important commercial examples of the dicy- lic monoterpenes, *see* CAMPHOR; PINENE.

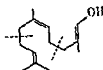
Tricyclic monoterpenes. These are not numerous. When camphorhydrazone is treated with yellow mercuric oxide in alkaline solution, tricyclene is formed.



Tricyclene

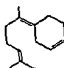
Sesquiterpenes. Few of the naturally occurring sesquiterpenes have been satisfactorily characterized. All of them contain three potential isoprene units.

Acyclic sesquiterpenes. Farnesol is a primary alcohol widely distributed in essential oils. Appreciable amounts have been found in ambrette seed oil.



Farnesol, $C_{15}H_{26}O$

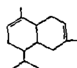
It is a valuable fixative for perfumes. Its spatial configuration has not been resolved.



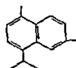
Zingiberene, $C_{15}H_{24}$

Monocyclic sesquiterpenes. Zingiberene is the principal constituent of ginger oil.

Dicyclic sesquiterpenes. Cadinene is widely distributed in nature and forms the main ingredient



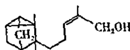
Cadinene, $C_{15}H_{24}$



Cadalene, $C_{15}H_{18}$

of oil of cubebs. Dehydrogenation yields cadalene, a substituted naphthalene.

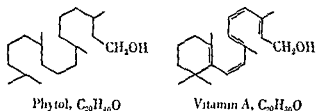
Tricyclic sesquiterpenes. Santalol occurs in sandalwood oil and is used in perfumery The esters



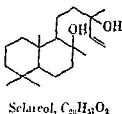
α Santalol

are also used in perfumes and once had some importance in medicine.

Diterpenes. Of the diterpenes theoretically composed of four isoprene units, the structures of such compounds as rosin, sclareol, vitamin A, and the chlorophyll alcohol, phytol, have been determined.



Phytol, $C_{20}H_{40}O$ Vitamin A, $C_{20}H_{30}O$

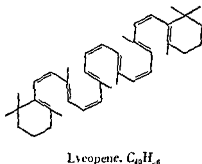


Sclareol, $C_{20}H_{36}O_2$

Vitamin A is commercially produced using *β*-ionone as a starting material

Triterpenes. For a discussion of this important class of terpenes, *see* TRITERPENE.

Tetraterpenes. The best-known members of the tetraterpenes belong to the carotenoid pigments from plants and animals. They are colored materials distinguished by numerous conjugated olefinic bonds.



Lycopene, $C_{40}H_{56}$

Synthetic linalool is the starting compound for the synthesis of lycopene, the red coloring matter in tomatoes

Polyterpenes. Various rubbers from *Hevea*, *Guayule*, and other plants yield small amounts of isoprene on pyrolysis. Their molecules are composed of several thousand of the C_5 units *See* ESSENTIAL OILS; ISOPRENE; MENTHOL; ROSIN, TURPENTINE. [E L S]

Bibliography: J. L. Simonsen (ed.), *Terpenes*, 5 vols., 1947——.

Terra cotta

Burned or fired clay (baked earth) modeled or molded into objects. When the material is unglazed, it is sometimes called biscuit and is of a reddish color and moderately porous. When it is glazed, it is used in items as varied as roof tiles and Tanagra figurines.

Selected clays are ground, mixed with grog (preground and burned clay) and water in varying proportions, molded or modeled into the desired shape, and then fired in a kiln. Terra cotta is widely used in making pottery and small sculpture and in other art modeling and similar handicrafts. Machine fabrications are used in building and architectural applications such as unglazed tiles for rough interior construction and for fireproofing, glazed exterior or interior wall and roof tiles, and decorated and glazed tiles, as building ornamentation. See CLAY PRODUCTS, ARCHITECTURAL. [G.CO]

Terracing (agricultural)

A method of shaping land to control water erosion on steep slopes used for cropping and other purposes. In early practice the land was shaped into a series of nearly level benches or steplike formations. Modern practice in terracing, however, consists of the construction of low-graded channels or levees to convey the excess rainfall from the land at nonerosive velocities. The physical principle involved is that when water is spread in a shallow stream, its flow is retarded by the roughness of the bottom of the channel and its carrying, or erosive power, is reduced In areas of low rainfall and absorbent soils, nearly level terraces are used Since direct impact of rainfall on bare land churns up the soil and the stirring effect keeps it in suspension in overland flow and rills, terracing does not



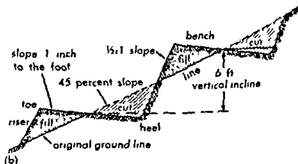


Fig 1 (a) True bench terraces consist of a flat step, or bench, for cultivation (notice the furrows) and a slope covered with grass (b) Cross section of an excavated bench terrace Sketch shows bench terrace constructed on a 45% slope, using a vertical interval of 6 ft, and having a back slope of ½:1 on the riser The bench is approximately 10 5 ft wide with an in slope of 1 in /ft of width (From USDA, Soil Conserv. Service, A Manual on Conservation of Soil and Water, Handbook for Professional Agricultural Workers, Agriculture Handbook No 61, June, 1954)

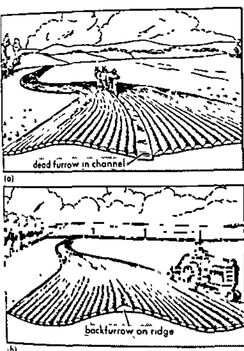


Fig. 2 Maintaining terraces by plowing. (a) The channel of the channel terrace can be enlarged by plowing it out. Between channels, turn as many of the furrows uphill as possible to offset the natural soil movement down the slope (b) The ridge of the ridge terrace can be enlarged by backfurrowing to it The location of the dead furrow should be changed from year to year to avoid excessive depression at any one point (*From USDA, Soil Conserv. Service, A Manual on Conservation of Soil and Water, Handbook for Professional Agricultural Workers, Agriculture Handbook No 61, June, 1954*)

prevent sheet erosion. It serves only to prevent destruction of agricultural land by gullying and must be supplemented by other erosion-control practices, such as grass rotations, cover crops, mulching, contour farming, strip cropping, and increased organic matter content. All these give the soil better structure and increased absorptive capacity. *See* EROSION, SOIL CONSERVATION

Types of terraces. There are two distinctive types of terraces, the bench and the ridge, or channel, type (Figs. 1 2) The be[illegible]

[illegible] cultivation, orchards or vineyards, or for landscaping. In old agricultural areas of the Philippines, Asia, the Near East, and South America, the risers were commonly of stone and many are still in use after hundreds of years. In southeastern United States, benches have been developed on slopes with a rise as high as 30 ft per hundred feet by careful management of the vegetative risers. However, in the last 10–15 years those have been largely abandoned for general farming because they are not adapted to the efficient use of large power equipment.

Modern tractor farming greatly increased the need for erosion control but the tractor also furnished the power for increased terracing. The early American practice of constructing hillside ditches across the slopes of fields to prevent up- and downhill gullying was followed by the development of the more easily controlled levee or ridge. This was called a narrow-based ridge terrace and was in some cases vegetated and developed into a bench-type terrace. By plowing to this narrow levee and maintaining a drainage channel, its base may be widened. It is then called the broad-base terrace. Difficulty in maintaining a water channel above the level of the surface of the soil resulted in the development of a ditch-type terrace channel, notably in the deep South where intense rains frequently broke the levee types.

All the types of erosion control described above are still in use, but the ditch, or channel, type is most extensively used because of low cost of maintenance and safety against breakage. The channel is given an increased slope as it goes down the hill in order to increase its carrying capacity for the accumulating water of its enlarged drainage area Since forces of erosion are constantly at work filling the channel with erosional debris, maintenance is of prime importance and channels must be plowed out or otherwise kept open as an essential part of the management of the land.

Outlet construction. Since terrace channels concentrate rainfall on hillsides, outlets are a major feature of any successful terrace system. There are many different schemes for outlet construction, utilizing masonry structures such as storm sewers, concrete flumes or drop inlets on steep land, and vegetation such as grass or other thick growing crops on gentle slopes Modern practice in outlet construction calls for the use of natural channels for outlets and careful shaping and vegetating before large concentrations of water are turned into them

Land farming. The practice of terracing has been materially changed by the development of large earth moving equipment and the necessity for efficient use of manpower in production Modern practice, commonly called land-forming or land-shaping, consists of reforming the land sufficiently to construct uniformly spaced parallel terraces and thus eliminate short or uneven "point rows" between the terraces This enables the use of large equipment, such as four- or six-row planters and cultivators, and avoids unnecessary turning on the terraces On gently sloping land this practice is combined with filling the low spots in the field with soil from the terrace channel and smoothing and leveling the entire crop area to give uniform surface drainage with no puddles or wet areas remaining to reduce yields The practice of land-forming is gradually being extended onto steeper lands where the subsoils are sufficiently fertile or where the surface soil is of sufficient depth to per-



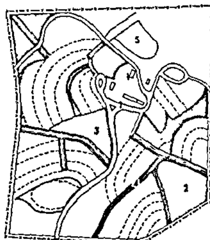


Fig. 3. Airplane view and diagram of layout of terracing system for land averaging 6% slope. Four-row production equipment is used exclusively, leaving irregular patches in grass to avoid short rows

mit the cuts that are necessary without serious effects on yield. Lands with less uniform slope may be laid off with terraces at variable horizontal spacings with variable graded channels, and the crops then planted in long rows The remaining irregular unplanted areas are planted to erosion-resisting vegetation (Fig. 3). Land with 5- to 10-ft fall per hundred feet may be farmed in this fashion with large and efficient machinery. The uniformity of the slope of the land determines the upper limit of practicability of this practice. *See* AGRICULTURAL MACHINERY; AGRICULTURAL SOIL AND CROP PRACTICES; AGRICULTURE (DRAINAGE) [M.L.N.]

Bibliography: *See* AGRICULTURAL SCIENCE (PLANT)

Terrain areas, world-wide

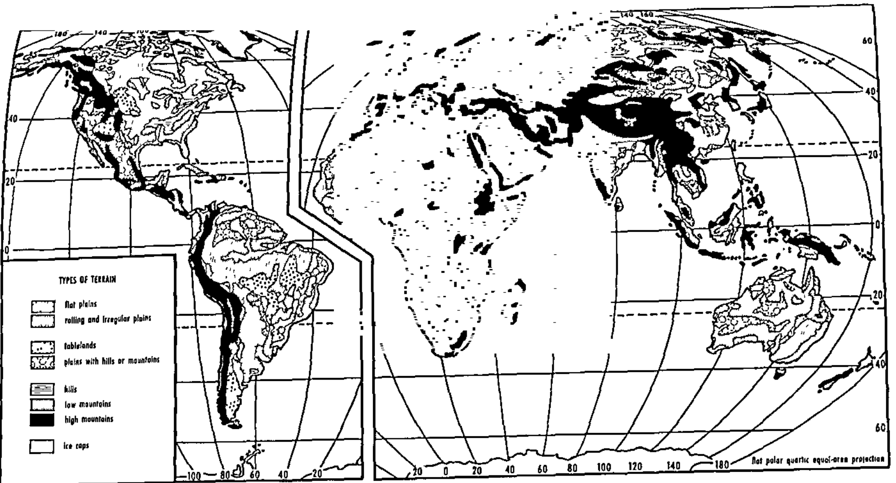
Subdivisions of the continental surfaces distinguished from one another on the basis of the form, roughness, and surface composition of the land. These areas of distinctive landforms are the product of various combinations and sequences of events involving both deformation of the earth's crust and surficial erosion and deposition by water, ice, gravity, and wind. The pattern of landform differences is strongly reflected in the arrangement of such other features of the natural environment as climate, soils, and vegetation These regional associations must be carefully reckoned with by man in his planning of activities as diverse as agriculture, transportation, city development, and military operations.

The accompanying map distinguishes among eight classes of terrain, on the basis of steepness of slopes, local relief (the maximum local difference in elevation), cross-sectional form of valleys and divides, and nature of the surface material. Approximate definitions of terms used and percentage figures indicating fraction of land area occupied by each class are as follows: (1) flat plains—nearly level land, slight relief, 6%; (2) rolling and irregular plains—mostly gently sloping, low relief, 30%; (3) tablelands—upland plains cut at intervals by deep valleys or canyons, moderate to high relief, 5%; (4) plains with hills or mountains—plains surmounted at intervals by hills or mountains of limited extent, 13%; (5) hills—mostly moderate to steeply sloping land of low to moderate relief, 8%, (6) low mountains—mostly steeply sloping, high relief, 14%; (7) high mountains—mostly steeply sloping, very high relief, 14%; (8) ice caps—surface material, mostly glacier ice, 10%.

The continents differ considerably. Australia, the smoothest continent, has only one-fifth of its area occupied by hill and mountain terrain as against one-third of North America and more than one-half of Eurasia Antarctica is largely ice covered; the only other great ice cap is on Greenland.

North America, South America, and Eurasia are alike in that most of their major mountain systems are linked together in extensive cordilleran belts. These form a broken ring about the Pacific basin, with an additional arm extending westward across southern Asia and Europe. The principal plains of Asia and Europe lie on the Atlantic and Arctic sides of the cordilleras, but are in part separated from the Atlantic by lesser areas of rough terrain.

Most of Africa and Australia, together with the eastern uplands of South America and the peninsulas of Arabia and India, show great similarity to one another They lack true cordilleran belts, and are composed largely of upland plains and tablelands, locally surmounted by groups of hills and mountains, and in many places descending to the sea in rough, dissected escarpments. *See* CORDILLERAN BELT; HILL AND MOUNTAIN TERRAIN; PLAINS [E.H HA.]



Distribution of terrain classes over the earth. The subdivisions are intended to bring out only the most striking contrasts among land surfaces. Percentage of area occupied by each class is given in accompanying discussion (*From V. C Finch, G. T. Trewartha, A. H. Robinson, and E. H. Hammond, Elements of Geography, 4th ed , 1957*)

Terrain-clearance indicator

A landing altimeter using radio waves to determine the height of the aircraft in relation to the terrain directly below it and capable of precision measurements to very low altitudes.

Low approach is defined as a descent from an initial altitude to a point near the ground. From the time that the aircraft reaches this point near the ground to the time that the wheels of the aircraft actually contact the ground, the pilot's chief interest is in his proximity to the ground and the speed with which he is approaching it. It is apparent, therefore, that means for measuring altitude and the rate of change of this altitude must be employed. This device is the landing altimeter, or terrain-clearance indicator, which may have a maximum altitude capability not greatly in excess of 200 ft. Some, however, have dual scales capable of reading to either 5000 or 500 ft.

To obtain the necessary precision all of these altimeters are of the frequency-modulated type. The frequency of the transmitter is changed linearly with time following a saw-toothed pattern. Early developments operated at a frequency of about 400 megacycles (Mc) but the newer designs operate at 4000 Mc and the frequency bands are swept at the rate of 100 Mc/sec. The radiator usually consists of an electromagnetic horn mounted flush with the surface of the aircraft. A second horn is employed as receiving antenna. In the receiver the signal reflected from the ground is mixed with a signal at transmitter frequency, producing a beat-note in the input to the altimeter receiver. Since the transmitted frequency is constantly being changed linearly with time, the frequency difference is a measurement of the time the transmitted energy was en route to the ground and back. Therefore, the frequency of the beat-note is a direct indication of the altitude. The beat-note frequency is determined by suitable circuits and indicated on a direct-reading meter. The first derivative of the frequency with time is also determined to give rate of descent. Electrical outputs proportional to frequency and rate of change of frequency are then mixed with the glide-slope indication of the fixed-beam low-approach system and used to actuate the horizontal needle on the cross-pointer instrument or to actuate the automatic pilot. See ALTIMETER, RADIO; INSTRUMENT LANDING SYSTEM (ILS) [P.C.S.]

Bibliography: P. C. Sandretto, *Electronic Avigation Engineering*, 1958.

Terrapin

A brackish-water, aquatic turtle of the genus *Malaclemys*, family Emydidae, related to the pond and box turtles. There are six forms, all found in the coastal waters of the eastern United States from New England to Texas. Some authorities recognize two species; others place all forms in the species *M. terrapin* and recognize six subspecies.

This turtle was formerly an important food animal along the central Atlantic Coast but is now scarce. It is moderately small, with a carapa length of 7½ in. or less, the females being som what larger than the males. The diamond ba terrapin has a keeled, depressed carapace, t plates of which are marked with prominent co centric ridges. Terrapins have been somewhat stored by a Federal hatchery, and they have be introduced on the West Coast. See REPTILI TURTLE. [J.D.]

Terrestrial ecosystem

A term that distinguishes the complex of ecosy tems of the land surfaces of the earth from fres water and marine ecosystems. It encompasses ec systems that exist on the continents and islands the world and comprehends a series of dynam open interaction systems that include living form (animals, plants, and microorganisms) and the nonliving environment (soils, geological formation and atmospheric constituents) and the activitie interrelations, chemical reactions, physical change and all other phenomena of each. Energy that e ters these systems, chiefly in the form of sunligh circulates through the systems and powers the li processes of the organisms, influences the rate an nature of chemical reactions and physical change and is partially accumulated in the bodies of orga isms and in other chemical and physical state See ECOSYSTEM; FRESH-WATER ECOSYSTEM; MARIN ECOSYSTEM.

Comparison of ecosystems. Terrestrial ecosy tems differ from aquatic ecosystems in several im portant respects. The most obvious difference in th abiotic components of the systems is the basi physical contrast between the media. Terrestrial or ganisms are surrounded by air, a mixture comprise of gaseous elements and compounds. Water vapo is present in the atmosphere but forms only small portion of the total volume of the air.

Liquid water, the medium of the aquatic envi ronment, is much more dense, less fluid, less trans parent, and has a much greater thermal stability than air. The aquatic environment is marked by a superabundance of moisture, a lack of temperature extremes, slow changes in temperature, the disper sion of most of the essential mineral elements in the engulfing medium (usually in low concentra tion), a relatively short supply of oxygen and a high content of carbon dioxide, and a sharp de crease in illumination with slight increase in depth. On the other hand, in the terrestrial environment water is often in critically short supply, extremes and rapid changes of temperature are common, mineral elements are limited in occurrence to the substrate (generally in relatively high concentra tions), and light is intense, at least at upper levels of the vegetation.

Adaptations to the habitat. The contrast in density, composition, and physical properties of the media in terrestrial and aquatic habitats is accompanied by and is either responsible for, contributory to, or correlated with, contrasts in the geomorphologic development of the habitat, climatological

features of the environment, pedological development, and morphologic and physiologic characteristics of the biota. Many organisms that live in the water have not developed, or have lost, devices that afford protection against water loss; indeed, many are able to absorb moisture and nutrients through the entire surface of their bodies. Many terrestrial organisms have evolved impervious cuticular coatings, surficial cuticular coverings, behavioral adaptations, and physiologic adaptations that reduce water loss or make exceptionally efficient use of available water. Land plants, particularly the ferns and seed plants, have developed various water- and nutrient-collecting structures like rhizoids and roots. Correlated with these is the development of a complex translocation system through which water and dissolved minerals move from root to leaves and food circulates through the plant. Water plants generally have poorly developed roots and translocation systems.

Structural adaptations. Particularly in animals, the body shape of terrestrial species generally presents a small surface area per unit of volume, whereas marine animals, especially the mobile forms, are often elongate or flattened and hence present a large amount of surface per unit of volume. To some extent, body shape in motile forms is correlated with the density of the media and represents a streamlining adaptation that facilitates the movement of aquatic animals. Another structural

such development in aquatic plants, which are

worts, fungi, and lichens are most characteristic and abundantly represented on land, while water is the chief domain of the algae. Thus the diversity in gross structure and life form, as well as in species, is much greater on land. A tremendous variety of forests, grasslands, savannas, scrubs, succulent deserts, tundras, and other forms of vegetation characterize the terrestrial environment (see CLIMAX PLANT FORMATIONS). The range of physiognomy of aquatic vegetation is much less and stratification or layering is absent or weakly developed in aquatic communities.

Habitat distribution of organisms. Warm-blooded animals, snakes, and lizards are typically land dwellers. Insects are abundant and ecologically important on land, less so in fresh water, and are virtually absent from salt-water habitats. Fish, bivalves, and various lower animals are limited to or at least most abundant in aquatic environments. The bulk of aquatic vegetation, particularly where the water is deep, is floating rather than rooted while that of land is almost entirely rooted and stationary. In contrast, many aquatic animals are immobile and depend on water currents to deliver food to their vicinity; while most terrestrial animals are mobile. Land animals are generally capable of much more rapid movement in their dispersed medium than are comparable aquatic animals, which occupy a very dense medium.

Extent of terrestrial ecosystem. Although the aerial environment is well lighted from the upper limits of the atmosphere to the ground, the thickness of the terrestrial ecosystem is limited by two sets of circumstances. First the vertical development of land vegetation is restricted by the characteristics and limitations of supporting tissues, by the periodic stresses and breakage produced by strong winds, by the limitations of water-raising [illegible] Secondly, the

oxygen becomes a limiting factor, and water may saturate the soil at a depth of a few inches or feet or it may be in critically short supply. Thus the tallest living tree is a eucalyptus 364 ft high, and the roots of only a few species of plants extend to a depth of 100 ft. Discounting airborne spores, insects, and similar material accidentally carried into the upper atmosphere, the maximum thickness of the terrestrial ecosystem in a given locality is less than 400 ft. Usually, it does not exceed 150 ft in thickness in forested areas and is much thinner in areas that support only grassland or tundra. On the other hand, the inhabited zone of a marine environment may be greater than 30,000 ft thick, although food-producing plants are limited to the surface layers due to limits of light penetration.

Temperature. The great seasonal temperature variations and rapid temperature changes in most terrestrial regions force periods of dormancy in plants and poikilothermic animals whose body temperatures approximate the environment. Homoiothermic or warm-blooded animals, on the other hand, can remain active even during very cold weather if sufficient food is available. Aquatic plants and animals are not exposed to rapid thermal changes and the amplitude of the seasonal thermal cycle is much less in the water than in most terrestrial habitats. Not only is temperature regulation a greater problem in terrestrial habitats, but the functioning of the entire terrestrial ecosystem is controlled to a much greater extent by temperature changes. The fluidity of the atmosphere, the cyclonic circulation of huge air masses, and the variations in solar energy result in considerable diurnal thermal variations as well as seasonal and secular variations. During the growing season, or nondormant periods, these thermal variations impose corresponding variations in the rate of productivity of the green plants (through their effect on the rate of photosynthesis), on the rate of herbivore and carnivore harvesting activity (through their effect on the physiology of the organisms; especially as they impose dormancy due to low temperatures or estivation due to high temperatures), and on the energy expenditures of homoiothermic animals (through their effect on the temperature gradient

between the air and the bodies of the animals and the respiratory activity required to maintain body temperature). See HIBERNATION; HOMEOSTASIS.

Seasonal effects. During the cold or dry dormant seasons, the primary production of the terrestrial ecosystem is drastically reduced or even halted. Most herbivorous animals are inactive. The activities of herbivores that remain operative and the activities of saprophytes and scavengers that continue to feed during the unfavorable seasons result in a net decrease in the standing crop, for they feed on materials elaborated during the preceding favorable season or seasons. Most poikilothermic carnivores are forced to remain dormant or die during cold seasons, but homoiothermic carnivores may remain active. In the latter instance, the carnivores' food preferences may change during the cold season, or their methods of hunting may be changed to compensate for the cold-season habits of their prey. Migrations of populations are related closely to seasonal changes. Migration of some birds, for example, has been found to be triggered by changes in daylength But avian populations that are dependent on insects migrate to and from an area at times that correspond closely with the vernal increase and autumnal decrease in insect populations. Other species that are dependent on plant foods migrate at times when the daylight period becomes too short to allow the consumption of sufficient plant food to carry the animals through the longer dark period Many grazing and browsing mammals migrate vertically in mountainous areas or horizontally in level areas to obtain food and to benefit from more favorable climatic conditions See MIGRATORY BEHAVIOR

Unfavorable seasons also affect aquatic communities, especially fresh-water communities, but the seasonal changes generally are less intense, the activity of primary producers generally does not ebb as low as in terrestrial communities, and the limiting factors are most often a lack of oxygen and nutrients rather than extremely low temperatures or the unavailability of water. Animal activity in aquatic habitats is also reduced during unfavorable seasons, but the level of minimum activity in the aquatic environment of deeper fresh-water bodies and in the oceans is considerably higher than that in the terrestrial environment

Catastrophic agents. Fire, windthrow, and many other catastrophic agents are peculiar to the terrestrial environment and play important roles in the functioning of the ecosystem. Furthermore, the terrestrial environment is the habitat of man and is more subject to human interference than the aquatic environment, particularly the marine segment of it. Human influence is so great that man has substituted artificially maintained ecosystems for natural ecosystems. The artificial ecosystems are represented on land by farmed areas, managed forests, urbanized areas, and similar developments See ECOLOGY, APPLIED.

Energy. Terrestrial ecosystems differ from aquatic ecosystems in another significant aspect. They are dependent on stored nutrients that m have been residual for hundreds of thousands years. They are marked by a continual net loss nutrients and other physical components throu erosion and leaching, while there is a concomita net gain to the aquatic ecosystems. Except in ce tain restricted areas, the use of aquatic food terrestrial animals, the phenomenon of salt spra the filling of lakes and bogs, and the harvest of m rine and fresh-water organisms by human activiti represent a small return of materials to the terre trial environment. Much larger returns are ma by crustal upheavals that expose sections of t ocean floor and once again make available to te restrial organisms the materials accumulated on t ocean bottom The reverse also takes place, a former land areas may be depressed beneath t surface of the oceans. These crustal movements r sult in a long-term recirculation of mineral el ments as well as a periodic renewal of erosion c cles. The terrestrial and aquatic ecosystems a also connected in a great many other direct wa so that it is most logical to consider the enti earth and its envelope of gases to be the ultima ecosystem. However, this world ecosystem is open system and is dependent upon an external su ply of energy, chiefly from the sun, and is infl enced by a variety of forces of external origin.

Trophic levels. If the weight (biomass) of t living components of an ecosystem at a given m ment is determined, separated into weights for t various trophic levels (green plants, herbivore predators, scavengers, and saprophytes), and fi ured graphically, a pyramidal form of graph r sults. In the terrestrial ecosystem, the green plan are relatively long lived; some live for sever weeks, others for years, a few for centuries. B cause the green vegetation is the primary produce level and thus limits the bulk of organisms th can feed upon it and because the green vegetatio is long lived, its biomass is greater than that any other trophic level or of all other trophic leve combined In aquatic ecosystems, however, this not necessarily the case Where short-lived plankto forms are the predominant green vegetation an longer-lived fish and bottom organisms are the he bivores and carnivores, the biomasses of the highe trophic levels often exceed that of the primary pr ducers Of course, when the annual production o the various trophic levels in either ecosystem is cal culated, the production of the green vegetation i greatest.

Productivity From the meager quantitative dat available on the primary productivity of variou ecosystems, it appears that no generalization ca be made concerning the comparative productivit of terrestrial and aquatic communities. Further, i is apparent that productivity in either environmen is not a regional characteristic, but is dependen wholly upon the ecosystem involved. According t summaries made by E Odum (1953), an Ohio corn field was found to produce 862 metric tons of or ganic carbon per square ilo eter per year, the

in a New York apple orchard produce 526, pean forests produce an average of 225, cultiland on the average produces 160, a dry grass- produces 48, and a desert produces only 6 tons. H Odum, et al. (1958) present data indicate that a montane rainforest in Puerto produces approximately 1100 metric tons, and presented by A Krogh (1934) indicate that Danish beech forest may produce 1500–2000 metric tons.

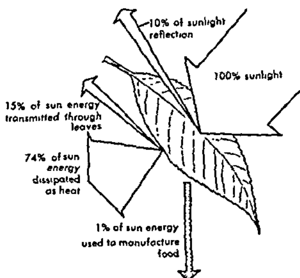
For comparison, data cited by E. Odum (1953) that several lakes in the north-central United States average 111–480 metric tons per square kilometer per year and that several marine communities average 60–1000 metric tons. Significantly higher annual primary production has been reported by Odum (1957) for a fresh-water spring in Florida, 6390 metric tons, and by E. Odum and H. Odum (1955) for a coral atoll in the Pacific Ocean, 8328 metric tons

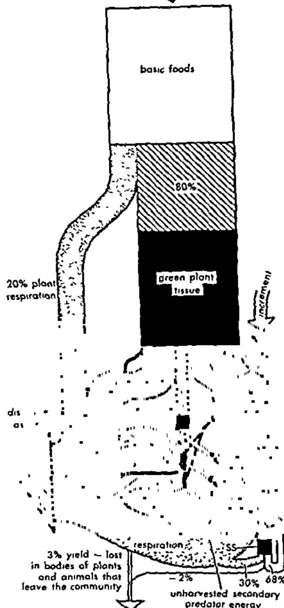
Trophic structure. The trophic structure of terrestrial and aquatic ecosystems is strikingly similar (see table). There is more variability within the types of analyzed terrestrial communities than there is between those communities and the aquatic types. In addition, the relative efficiency of the vegetation, in terms of the proportion of incident solar radiation converted to stored form in organic carbon compounds, is virtually identical for the terrestrial and fresh water communities listed, although efficiency of the tropical atoll community was considerably greater. In all the ecosystems, a rela-

United States have been estimated by J. S McCormick (1959) and are an example of terrestrial systems in general (see illustration). The energy used each year by the organisms in an acre of forest is approximately equivalent to the electricity required to supply an average New York household with power for nearly half a century. Virtually all the energy that enters the system directly is in the form of sunlight, but only about 1% of the available solar energy is actually transformed by green plants into the chemical energy of food. Approximately 10% of the solar energy is reflected from the plant surfaces, 15% is passed through the leaves, and 74% is dissipated as heat.

A portion of the energy stored in basic foods or

... the energy of plant tissues harvested and utilized by herbivores such as insects, rodents, and ... is dissipated as heat by respiration, and part





Pathways and magnitude of energy transfers in a beech-maple forest in the north-central United States (From J. S. McCormick, The Living Forest, Harper, 1959)

between the air and the bodies of the animals and the respiratory activity required to maintain body temperature). *See* HIBERNATION; HOMEOSTASIS.

Seasonal effects. During the cold or dry dormant seasons, the primary production of the terrestrial ecosystem is drastically reduced or even halted. Most herbivorous animals are inactive. The activities of herbivores that remain operative and the activities of saprophytes and scavengers that continue to feed during the unfavorable seasons result in a net decrease in the standing crop, for they feed on materials elaborated during the preceding favorable season or seasons Most poikilothermic carnivores are forced to remain dormant or die during cold seasons, but homoiothermic carnivores may remain active. In the latter instance, the carnivores' food preferences may change during the cold season, or their methods of hunting may be changed to compensate for the cold-season habits of their prey. Migrations of populations are related closely to seasonal changes Migration of some birds, for example, has been found to be triggered by changes in daylength But avian populations that are dependent on insects migrate to and from an area at times that correspond closely with the vernal increase and autumnal decrease in insect populations. Other species that are dependent on plant foods migrate at times when the daylight period becomes too short to allow the consumption of sufficient plant food to carry the animals through the longer dark period Many grazing and browsing mammals migrate vertically in mountainous areas or horizontally in level areas to obtain food and to benefit from more favorable climatic conditions. *See* MIGRATORY BEHAVIOR

Unfavorable seasons also affect aquatic communities, especially fresh-water communities, but the seasonal changes generally are less intense, the activity of primary producers generally does not ebb as low as in terrestrial communities, and the limiting factors are most often a lack of oxygen and nutrients rather than extremely low temperatures or the unavailability of water. Animal activity in aquatic habitats is also reduced during unfavorable seasons, but the level of minimum activity in the aquatic environment of deeper fresh-water bodies and in the oceans is considerably higher than that in the terrestrial environment.

Catastrophic agents. Fire, windthrow, and many other catastrophic agents are peculiar to the terrestrial environment and play important roles in the functioning of the ecosystem. Furthermore, the terrestrial environment is the habitat of man and is more subject to human interference than the aquatic environment, particularly the marine segment of it. Human influence is so great that man has substituted artificially maintained ecosystems for natural ecosystems. The artificial ecosystems are represented on land by farmed areas, managed forests, urbanized areas, and similar developments. *See* ECOLOGY, APPLIED.

Energy. Terrestrial ecosystems differ from aquatic ecosystems in another significant aspect. They are dependent on stored nutrients that may have been residual for hundreds of thousands of years. They are marked by a continual net loss of nutrients and other physical components through erosion and leaching, while there is a concomitant net gain to the aquatic ecosystems. Except in certain restricted areas, the use of aquatic food by terrestrial animals, the phenomenon of salt spray, the filling of lakes and bogs, and the harvest of marine and fresh-water organisms by human activities represent a small return of materials to the terrestrial environment. Much larger returns are made by crustal upheavals that expose sections of the ocean floor and once again make available to terrestrial organisms the materials accumulated on the ocean bottom. The reverse also takes place, and former land areas may be depressed beneath the surface of the oceans. These crustal movements result in a long-term recirculation of mineral elements as well as a periodic renewal of erosion cycles. The terrestrial and aquatic ecosystems are also connected in a great many other direct ways, so that it is most logical to consider the entire earth and its envelope of gases to be the ultimate ecosystem. However, this world ecosystem is an open system and is dependent upon an external supply of energy, chiefly from the sun, and is influenced by a variety of forces of external origin.

Trophic levels. If the weight (biomass) of the living components of an ecosystem at a given moment is determined, separated into weights for the various trophic levels (green plants, herbivores, predators, scavengers, and saprophytes), and figured graphically, a pyramidal form of graph results. In the terrestrial ecosystem, the green plants are relatively long lived; some live for several weeks, others for years, a few for centuries. Because the green vegetation is the primary producer level and thus limits the bulk of organisms that can feed upon it and because the green vegetation is long lived, its biomass is greater than that of any other trophic level or of all other trophic levels combined In aquatic ecosystems, however, this is not necessarily the case. Where short-lived plankton forms are the predominant green vegetation and [illegible] trophic levels often exceed that of the primary producers Of course, when the annual production of the various trophic levels in either ecosystem is calculated, the production of the green vegetation is greatest

Productivity From the meager quantitative data available on the primary productivity of various ecosystems, it appears that no generalization can be made concerning the comparative productivity of terrestrial and aquatic communities. Further, it is apparent that productivity in either environment is not a regional characteristic, but is dependent wholly upon the ecosystem involved According to summaries made by E. Odum (1953), an Ohio cornfield was found to produce 862 metric tons of organic carbon per square [illegible]eter per year, the

a New York apple orchard produce 526, ...n forests produce an average of 225, culti... ...nd on the average produces 160, a dry grass... ...oduces 48, and a desert produces only 6 tons. H Odum, et al. (1958) present data ...licate that a montane rainforest in Puerto ...oduces approximately 1100 metric tons, and ...esented by A. Krogh (1934) indicate that ...sh beech forest may produce 1500–2000 tons.

...l primary production has been reported by ...um (1957) for a fresh-water spring in Flor... ...390 metric tons, and by E. Odum and H. Odum ...) for a coral atoll in the Pacific Ocean, 8328 ...c tons.

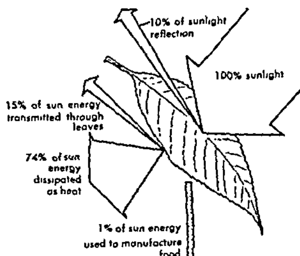
...phic structure. The trophic structure of ter... ...al and aquatic ecosystems is strikingly similar ...table) There is more variability within the ...s of analyzed terrestrial communities than ...is between those communities and the aquatic ...s. In addition, the relative efficiency of the veg... ...on, in terms of the proportion of incident solar ...ation converted to stored form in organic car... compounds, is virtually identical for the terres... ...l and fresh water communities listed, although ...efficiency of the tropical atoll community was ...siderably greater. In all the ecosystems, a rela... ...ly low proportion of available solar energy is ...lized directly by the biota.

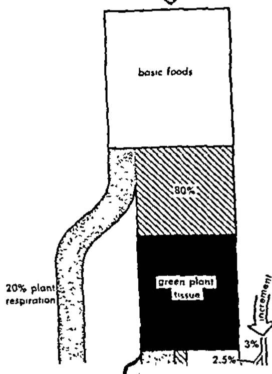
...Energy transfer. The energy transfers that oc... ...r in a beech maple forest community in the cen... ...l United States have been estimated by J. S. Mc...

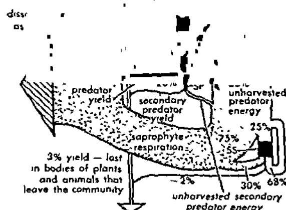
...ia forest is approximately equivalent to the elec... ...icity required to supply an average New York ...ity household with power for nearly half a cen... ...ury. Virtually all the energy that enters the sys... ...em directly is in the form of sunlight, but only ...bout 1% of the available solar energy is actually ...ransformed by green plants into the chemical en... ...rgy of food. Approximately 10% of the solar en... ...ergy is reflected from the plant surfaces, 15% is ...passed through the leaves, and 74% is dissipated as heat.

A portion of the energy stored in basic foods or ...

... of the energy of plant tissues harvested and ...tilized by herbivores such as insects, rodents, and ...deer is dissipated as heat by respiration, and part







...Pathways and magnitude of energy transfers in a beech-maple forest in the north-central United States. (From J. S. McCormick, The Living Forest, Harper, 1959.)

Trophic structure of various terrestrial and aquatic communities on the basis of biomass or energy-content of the standing crop

	Terrestrial communities			Fresh-water communities			Marine commu	
	Blue-grass meadow, Michigan	Beech forest, Den-mark	Montane rainforest, Puerto Rico	Cedar Bog Lake, Minne-sota	Webster's Lake, Wis-consin	Silver Spring, Flor-ida	Coral reef, Eniwetok Atoll	E gra No Se
Primary producers, %	78	93	98 4	89 3	86.8	94.3	83.1	79
Herbivores, %	16	7	1.6	9.0	10.2	4 3	15 6	19
Carnivores, %	6			1.6	3.4	1.5	1.3	0
Efficiency of primary producers*	1.2	1.0				1.2	5 8†	
Basis of figures	Energy	Biomass	Biomass	Energy	Biomass	B s	Biomass	Biom

is stored in the body tissues of the herbivores. The energy contained in the tissues of herbivores is utilized by predators and the tissues of predators are, in turn, utilized by secondary predators. At each step, some energy is lost to the community through respiration and part is unharvested or unassimilated The unharvested and unassimilated energy accumulates chiefly on the forest floor in the form of dead leaves, twigs, flowers, fruits, fallen trunks, dead bodies, feces, and liquid wastes and is utilized by scavengers and saprophytes. Ultimately, all the energy that enters the forest community is dissipated as heat by respiratory processes, is lost in the bodies of plants and animals that leave or are taken from the forest, or is lost in the form of heat evolved by forest fires. In natural communities, these losses are balanced, or at least offset, by an increment of materials that enter the community from other sources See ECOLOGY [J.S M.]

Bibliography: A. Krogh, Conditions of life in the ocean, *Ecol Monographs*, 4(3):421–429, 1934, J. McCormick, *The Living Forest*, 1959; E. P. Odum, *Fundamentals of Ecology*, 2d ed., 1959, H T. Odum, Trophic structure and productivity of Silver Springs, *Ecol. Monographs*, 27:55–112, 1957, H T. Odum, W Abbott, and R. Selander, Studies on the productivity of the lower montane rainforest of Puerto Rico (Abstract), *Bull. Ecol. Soc. Am.*, 39:85, 1958; H T. Odum and E P Odum, Trophic structure and productivity of a windward coral reef community on Eniwetok Atoll, *Ecol. Monographs*, 25(3):291–320, 1955.

Terrestrial electricity

Electric phenomena and properties of the earth. The term terrestrial electricity has been used in the broad sense to include atmospheric electricity Geoelectricity is perhaps the preferred term for denoting electric properties of the nonatmospheric earth and electrical phenomena which transpire in the earth. Earth currents is the term used by specialists to denote a class of natural electric phenomena which belong to a world-wide system. These phenomena are related to many of the changes of geomagnetism, and some of their aspects are correlated with disturbances of the ionosphere, with solar flares, with polar lights, and with t periodic variation of sunspots. Earth currents this restricted sense are the chief topic of th article; but a few other geoelectric phenomena a first briefly mentioned.

Local geoelectric phenomena. Electric curren which stray from sources of electric power (vaga bond currents, stray currents) have at times re quired the attention of engineers, but they ar mentioned here only because areas invaded by such currents cannot be used for measuring either earth currents or the variations of geomagnetism because the former are often intense as compared with the latter. The intensity, in amp/cm², of the natural earth currents is very small, generally less than 10^{-10} amp/cm². The potential gradient is so small that it is expressed in volts per kilometer. Other local phenomena or properties, such as the electrochemical condition of the soil, must be recognized and taken into account if confusion in the interpretation of earth-current data is to be avoided. One of these is the difference in the electrochemical property of different soils If in each of two such bodies of soil a probe (electrode) is embedded, the equivalent of a battery is formed; and when the two probes are connected by a conducting wire a current will usually flow through the wire. A pipeline laid in such soil conditions will start a current which helps to promote corrosion of the pipe unless the latter is insulated In both these cases there may be little if any current in the earth before the conductor is introduced.

Vertical earth currents This term is used to denote a phenomenon observed chiefly in mountains According to one interpretation electric currents flow in the earth from all sides of a mountain toward the top No plausible explanation for such a phenomenon has been proposed, and consequently it has the attraction of a mystery. It is known that when two probes with the connecting wire are set along the slope of a mountain, current usually flows in the wire from the lower toward the higher point Perhaps the soil between two different altitudes on a mountain slope contains some constituents of a battery, the latter being completed when the probes are installed, and current flows only

ı the earth potential measurements pertain. ecologists found that the active acidity (hy- :n ion concentration) of the soil on mountain :s increases as a nearly linear function of alti- This is attributed to the more rapid leaching . the soil of the alkaline constituents. Such hy- en ion data may be used to calculate the elec- l potential one might expect to measure in the opriate region if the probes used are equiva- to hydrogen electrodes. The results, using the ity data reported by S. A. Cain (1931) for a on in the Great Smoky Mountains, are: the ntial decreases with altitude at the rate of 0.2

entials made in the mountains of Europe. This ement may be enough to set up the explanation posed here as a target at which to aim future estigations of the puzzling vertical earth cur- ts.

Natural geoelectric currents. Such currents are uced when water moves across the earth's mag- ic lines of force. Evidence of such currents has n observed only in connection with tidal motion d some wave action. In the water the direction of : current is, for the Northern Hemisphere, from : left to the right of one who faces in the direc- n of the motion.

Discovery of earth currents. Soon after the first mmercial telegraph came into use (1844) there peared at times strange electric currents which truded on the telegraph lines and occasionally terfered with the sending of messages. Because a ngle line was used, which when in operation was nnected to earth at both ends, an additional hannel was offered for the electric currents of the rth. Close observation of the currents in lines of e British telegraph system led W. H. Barlow to

graph service.

The occasions when these currents are unusually ntense and impulsive are termed storms (earth- current storms) but they have no noticeable con- nection with weather. Other periods which are undisturbed are qualified as quiet or calm. These two different aspects of earth currents are dis- cussed separately after a brief description of the method of measurement.

Earth current measurement. No method has been devised to measure the earth-current density i directly; but i may be estimated from the relation $i = \varepsilon/r$, if r (the electric resistivity of earth) and ε (the gradient of electric potential) are known. Earth resistivity surveys have been made by a method which indicates to some extent the varia- tion of this property with depth, to depths of the order of 1 km, and also the variation with horizon- tal location. Such surveys have been made at four places where earth potentials have also been regis- tered, namely Watheroo, Western Australia; Ebro, Spain; Huancayo, Peru; and Tucson, Arizona. The data indicate that the over-all resistivity at a given place is essentially constant except for superficial decreases such as from dry to rainy periods, but that it may vary considerably from one site to an- other. For example, at the Watheroo Magnetic Observatory, Western Australia, 5000 ohm-cm may be taken as representative, whereas at the Ebro Observatory, Tortosa, Spain, 15,000 ohm-cm is ap- plicable. These values seem adequate for making a rough estimate of earth-current density when the gradient is determined. Because the appropriate resistivity does not vary appreciably with time, the temporal variations of the earth current density (the features of chief interest) will correspond with those of the gradient.

The potential gradient. This value is ascertained by the use of telegraph lines or with similar experi- mental lines constructed for the purpose. The potential difference between the points where the terminals of the lines make contact with the earth is usually registered automatically. This difference of potential divided by the straight line distance between the terminals gives what may be termed

the lines are so arranged that for one component the direction is north-south and for the other east- west. In this case the gradient is calculated, or con- structed graphically, just as in the case of the composition of forces; but when these directions do not differ by 90° a different treatment is required. This N-S, E-W arrangement also facilitates the comparison of results from different observatories. It is unimportant whether the course of the wire which connects the terminal points be direct or roundabout except that allowance for its resistance may be required.

Electrodes. Used at the terminal points to make contact with the earth, electrodes must be given the most careful attention if erroneous results are to be avoided. Nonpolarizable electrodes would seem to be ideal for such use, but they are im- practicable for permanent systems because of the high cost of installation and the considerable effort required for maintenance.

Lead, in the form of wire, has proved satisfactory as an electrode material. This is laid in trenches deep enough so that changes in temperature and moisture of the soil are insignificant at least during 1 day. A wheel-like array of trenches which has been used consists of one or two concentric circular trenches and four radial trenches. The lead wire is laid on each side of a trench. Such a pair is cross- connected at intervals and good connection is made at all points where radial wires cross those in the circular trenches. At the center of this array where the radial wires of the lead grid meet, a secure splice is made with the wire which con-

Trophic structure of various terrestrial and aquatic communities on the basis of biomass or energy-content of the standing crop

	Terrestrial communities			Fresh-water communities			Marine communities	
	Bluegrass meadow, Michigan	Beech forest, Denmark	Montane rainforest, Puerto Rico	Cedar Bog Lake, Minnesota	Webster's Lake, Wisconsin	Silver Spring, Florida	Coral reef, Eniwetok Atoll	Eel grass, North Sea
Primary producers, %	78	93	98.4	89.3	86.8	94.3	83.1	79.7
Herbivores, %	16	7 (herbivores and carnivores)	1.6 (herbivores and carnivores)	9.0	10.2	4.3	15.6	19.9
Carnivores, %	6			1.6	3.4	1.5	1.3	0.3
Efficiency of primary producers*	1.2	1.0				1.2	5.8†	
Basis of figures	Energy	Biomass	Biomass	Energy	Biomass	Biomass	Biomass	Biomass

* Ratio of energy transformation to incident solar energy.

† *Computed on basis of light reaching community rather than light reaching water surface.*

is stored in the body tissues of the herbivores. The energy contained in the tissues of herbivores is utilized by predators and the tissues of predators are, in turn, utilized by secondary predators. At each step, some energy is lost to the community through respiration and part is unharvested or unassimilated. The unharvested and unassimilated energy accumulates chiefly on the forest floor in the form of dead leaves, twigs, flowers, fruits, fallen trunks, dead bodies, feces, and liquid wastes and is utilized by scavengers and saprophytes. Ultimately, all the energy that enters the forest community is dissipated as heat by respiratory processes, is lost in the bodies of plants and animals that leave or are taken from the forest, or is lost in the form of heat evolved by forest fires. In natural communities, these losses are balanced, or at least offset, by an increment of materials that enter the community from other sources. *See* ECOLOGY. [J.S.M.]

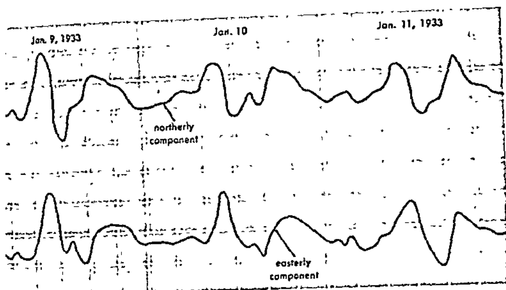
Bibliography: A. Krogh, Conditions of life in the ocean, *Ecol. Monographs*, 4(3):421–429, 1934; J. McCormick, *The Living Forest*, 1959; E. P. Odum, *Fundamentals of Ecology*, 2d ed., 1959, H. T. Odum, Trophic structure and productivity of Silver Springs, *Ecol. Monographs*, 27:55–112, 1957, H. T. Odum, W. Abbott, and R. Selander, Studies on the productivity of the lower montane rainforest of Puerto Rico (Abstract), *Bull. Ecol. Soc. Am.*, 39:85, 1958, H. T. Odum and E. P. Odum, Trophic structure and productivity of a windward coral reef community on Eniwetok Atoll, *Ecol. Monographs*, 25(3):291–320, 1955

Terrestrial electricity

Electric phenomena and properties of the earth. The term terrestrial electricity has been used in the broad sense to include atmospheric electricity. Geoelectricity is perhaps the preferred term for denoting electric properties of the nonatmospheric earth and electrical phenomena which transpire in the earth. Earth currents is the term used by specialists to denote a class of natural electric phenomena which belong to a world-wide system. These phenomena are related to many of the changes of geomagnetism, and some of their aspects are correlated with disturbances of the ionosphere, with solar flares, with polar lights, and with the periodic variation of sunspots. Earth currents in this restricted sense are the chief topic of this article; but a few other geoelectric phenomena are first briefly mentioned.

Local geoelectric phenomena. Electric currents which stray from sources of electric power (vagabond currents, stray currents) have at times required the attention of engineers, but they are mentioned here only because areas invaded by such currents cannot be used for measuring either earth currents or the variations of geomagnetism because the former are often intense as compared with the latter. The intensity, in amp/cm², of the natural earth currents is very small, generally less than 10^{-10} amp/cm². The potential gradient is so small that it is expressed in volts per kilometer. Other local phenomena or properties, such as the electrochemical condition of the soil, must be recognized and taken into account if confusion in the interpretation of earth-current data is to be avoided. One of these is the difference in the electrochemical property of different soils. If in each of two such bodies of soil a probe (electrode) is embedded, the equivalent of a battery is formed; and when the two probes are connected by a conducting wire, a current will usually flow through the wire. A pipeline laid in such soil conditions will start a current which helps to promote corrosion of the pipe unless the latter is insulated. In both these cases there may be little if any current in the earth before the conductor is introduced.

Vertical earth currents. This term is used to denote a phenomenon observed chiefly in mountains. According to one interpretation electric currents flow in the earth from all sides of a mountain toward the top. No plausible explanation for such a phenomenon has been proposed, and consequently it has the attraction of a mystery. It is known that when two probes with the connecting wire are set along the slope of a mountain, current usually flows in the wire from the lower toward the higher point. Perhaps the soil between two different altitudes on a mountain slope contains some constituents of a battery, the latter being completed when the probes are installed, and current flows only



…on hourly values of northerly and easterly …nt potential-gradients on three successive …quiet days at Tucson, Arizona.

…12 h the direction is south of SW and the …e 3.3 mv/km

…dographs for different places differ con… Nearly all have an elongated shape like …in Fig. 3; the direction of the major axis …upon location, and in some cases the figure …ed closely to the axis. In these cases the …is somehow restricted to flow almost only …is axis, but it changes from one direction …pposite usually twice each day. This limita… flow may be attributed to the way electri…nducting matter is distributed in the earth. …ter of the oceans is a much better conductor

…United States show this. Solar activity is …ated with the diurnal variation of earth cur… The amplitude increases as activity on the …measured by sunspot number, increases, but …is no marked change in the shape of the hodo…

…seasonal change of the diurnal variation is …that the hodographs are larger from spring …tumn than in winter, but at some places there …marked change in the shape of the hodographs …some months. The latter pattern applies es…ally to Tucson, Arizona. A plausible explana… for this will be given later. But another re…kable pattern feature appears there; after …ng shrunk in December, the hodographs are …denly blown up in January (with some excep…s), shrink again in February, and have a mark… shriveled appearance in March, but are well …anded again in April. Whether this January …maly depends upon conditions in the earth or …on a somewhat localized and seasonal feature …the ionosphere are still unanswered questions

Lunar diurnal variation. A lunar day depends upon the rotation of the earth with respect to the moon, just as the solar day depends upon the earth's rotation with respect to the sun. Because these

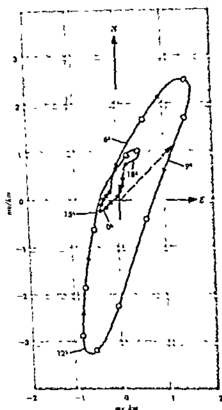


Fig. 3. Hodograph representing the potential-gradient vector of the average diurnal variation of earth currents at Tucson, Arizona, for 10 quiet days each month for 1932 to 1942, inclusive

nects the electrodes to the measuring instrument. This splice must be well insulated, protected from moisture, and placed at a depth where at least the diurnal variation of temperature is negligible.

Periodic variations; electrode potentials. Long-period variations of the electrochemical environ- … ion. … The … difference of contact potential between two electrodes would not be negligible if the aim is to detect steady unidirectional currents. These electrode potentials become a smaller part of the measured potential difference when the distance between electrodes is greater because the potential difference of the earth current is then greater. Distances as great as 300 km have been used, but satisfactory data for the diurnal variation and other variations of shorter period can be obtained with distances of 1 km. It is only such variations that are discussed in the rest of this article.

Earth-current storms. Some of these disturbances in earth currents are largely confined to the polar regions but others occur everywhere on the earth and at the same time Copies of the registrations for one such storm are reproduced in Fig. 1. This disturbance started shortly after 16 h Green-

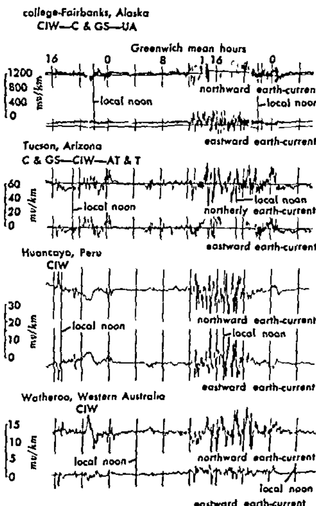


Fig. 1. Earth-current storm of April 30 to May 2, 1933, recorded at four places ranging in latitude from 65°N to 30°S showing events which occur simultaneously over the entire earth. (After a diagram in J. Wash. Acad. Sci., 26(7).276, 1936)

… for about 11 hours relative quiet prevailed; after this a more intense phase of the storm began and lasted until about 2 h, May 2. A correspondence between some of the details in these records can be traced but it is sufficient here to note the correspondence in general features. There is a similar correspondence between these records and those of the earth's magnetism for the same period. Apparently a relation between the disturbances in earth currents and those in geomagnetism exists. Other evidence bears this out. Both earth-current storms and magnetic storms are associated with the occurrence of auroras; both occur more frequently during periods when sunspots are larger and more numerous; for both there is a likelihood that 27 days after a storm another storm will occur (27 days is the rotation period of the sun).

It is to be expected that when magnetic changes are occurring electric current would be induced in the earth. The intensity of the earth current would be greater the more rapid the magnetic change. Slow magnetic variations which in time produce a larger change would induce only weak currents. Some evidence of such a relation is found when the electric and magnetic records are compared. But it also appears that sometimes the magnetic variation is caused by the variation of the earth current. It is conceivable that the latter relation may be noted at a place which is remote from the region to which the varying magnetic field is chiefly restricted, but within that region the former relation should be dominant

Solar diurnal variation. This is the most conspicuous feature of quiet or calm days. The earth currents then change in a fairly regular manner during the day. As can be seen in Fig 2, each component of the potential gradient of earth currents has two principal maxima and two principal minima. In the actual records there are irregular variations which do not appear here because the mean value for each hour was used in plotting these graphs Most of the smaller fluctuations shown on the graphs are eliminated in the average diurnal variation for a month or for a year. There are seasonal changes in this diurnal variation and also changes which are closely correlated with the sunspot number. These are most readily described with the aid of a hodograph of the potential gradient vector, such as that shown in Fig. 3. One vector only is shown here as a broken line extending from the origin of coordinates to a point on the hodograph It represents the average gradient vector at shortly before 9 A M. at Tucson, Arizona. Its direction is NE and its magnitude is 1.6 mv/km. The … than in the night. At about 7.5 h the direction is north of NE and its magnitude is about 2.8 mv/km;

nmer to winter. More current usually flows in
:m in summer than in winter and apparently
y shift farther north in summer. *This seems to*
indicated by the change in the character of the
lographs of the diurnal variation from summer
winter at Tucson, Arizona. During the months
m May to August, inclusive, hodographs from
out 8 h to 16 h are very narrow, whereas in the
ir winter months they have a rotund figure. That

v an inspection of Fig. 4.

A number of features of earth currents have een described or mentioned in this article without ffering an explanation. Several examples may well e cited. That earth current storms are correlated ith sunspot number, with aspects of auroras, and ith some features of radio and the fact that they end to recur after 27 days are cases in point. There are also questions with respect to quiet day henomena, *for example:* (1) Why *should the* world wide system of electric eddies be fixed with respect to the sun? (2) Why should it be most intense on the daytime side of the earth? (3) Why should it shift with season? (4) Why should it vary with sunspot number? (5) Why is there a lunar diurnal variation?

The best answers or explanations to these queries that are now available were developed to account for variable aspects of geomagnetism, and these, insofar as space allows, are discussed in the article on geomagnetism The broad qualitative aspects of these explanations can be readily extended to earth currents. [O H G]

Bibliography: O. H Gish, Electrical messages from the Earth: their reception and interpretation, *J Wash. Acad. Sci.*, 26:267–289, 1936; O H. Gish, General description of the earth-current measuring system at the Watheroo Magnetic Observatory, *Terrestrial Magnetism and Atmospheric Elec.*, 28:89–108, 1923, O. H. Gish and W. J. Rooney, Measurements of resistivity of large masses of undisturbed earth, *Terrestrial Magnetism and Atmospheric Elec.*, 30:161–188, 1925; M S. Longuet-Higgens, M. E. Stern, and H Stommel, *The Electrical Field Induced by Ocean Currents and Waves, etc.*, Woods Hole Oceanographic Inst Contrib. 690, 1954; W. J. Rooney, *Earth Current Results at Tucson Magnetic Observatory, 1932–1942*, Carnegie Inst. Wash Publ. 175, 1949; W J Rooney, *Terrestrial Magnetism and Electricity*, Physics of the Earth, vol. 8, 1939

Terrestrial force fields

Fields of forces, emanating from the earth, and extending with changing patterns into space. These include gravity, magnetic, electric, electromagnetic, and thermal fields.

Gravity field. An essential part of the gravity field is caused by the Newtonian attraction of the earth's mass, a small part by the centrifugal force of the rotating earth around its polar axis The strength of the gravity field decreases—outside the earth—with the square of the distance from the earth's center. Because of this force field, for instance, the moon and artificial satellites are held in orbits around the earth. The force center is the center of the earth. *See* TERRESTRIAL GRAVITATION.

Magnetic field. The earth's magnetic field has two force centers: the magnetic North Pole and the magnetic South Pole. Both poles are about 18° *from the geographic poles,* but their location changes continuously. Ninety-four per cent of the magnetic force at the face of the earth comes from the earth's interior; the remaining 6 per cent, from external forces, including the fields of the sun and moon. *See* TERRESTRIAL MAGNETISM; *see also* AEROMAGNETIC SURVEYING, ROCK MAGNETISM.

Electric field. Electrical properties of the earth, especially attributes of current and conductivities, are becoming better known and understood. Most of the electrical character of the earth's interior is surveyed only through assumption, mathematical probing, and other indirect investigation. Measurement and knowledge of electrical currents and of variations in electrical conductivity in both the rock shell of the earth and in its atmosphere are resulting *in widespread practical applications, particularly* in geophysical exploration and prospecting for minerals *See* TERRESTRIAL ELECTRICITY; *see also* ATMOSPHERE; ATMOSPHERIC ELECTRICITY.

Electromagnetic force field. The electromagnetic field includes electric currents produced by *the electrochemical activity of rocks in the earth's* crust and earth electric currents induced by ionospheric currents which correlate with diurnal changes in the earth's magnetic field. The electromagnetic field is subject to a growing amount of investigation in the upper reaches of the atmosphere and in the transition to outer space. *See* ATMOSPHERE, EARTH INTERIOR, IONOSPHERE.

Thermal field. The earth's temperature increases with depth about 1°/100 ft As a result a strong thermal force field exists along the earth's radius. *See* EARTH INTERIOR; GEOCHEMISTRY

For additional information *see* AERONOMY; GEOPHYSICAL EXPLORATION. [W.A H]

Terrestrial frozen water

Seasonally or perennially frozen waters of the earth, exclusive of the atmosphere Water in the frozen or solid state is the hexagonally crystallized, birefringent mineral known as ice Terrestrial ice occurs in the form of temporary seasonal accretions during the cold months and *in the perennial* ice cover represented by glaciers, land-fast ice (ice shelves) and subsurface ground ice in permafrost regions

Glaciers, past and present. *Under present* climatic conditions the semipermanent terrestrial ice cover is essentially glacial. Glaciers cover approximately 10% of the world's land area (14,972,138 km² estimated by R. F Flint, 1957). Of this ice, 96% lies in Greenland (1,726 400 km²) and Antarctica (12,650,000 km²), leaving only 4% of

days *differ in length* it is possible to separate the component of the diurnal variation attributable to a lunar influence from that attributable to a solar *influence*. When this is done it is found that there is a lunar diurnal variation of double period with an amplitude about one-fifth that of the solar diurnal variation. A corresponding lunar diurnal *variation is found for* geomagnetism—another of the numerous instances where there is similarity between these two aspects of nature.

One of the theories which has been developed to account *for the diurnal variation* of geomagnetism is known as the *dynamo theory. It is the only* one which accounts for a lunar influence (*see* GEOMAGNETISM). *In a quantitative* formulation of this theory, formulas are *derived to enable* calculation of the components of the potential-gradient vector for the diurnal variation in earth currents. The agreement of the calculated with the observed results is good enough, in most respects, to promote confidence in this theory.

Large-scale patterns and summary. A world *view of the* systems of *electric* currents in the earth is shown in Fig. 4. This is a model which exactly represents the annual mean diurnal variation for Watheroo, Australia; Tucson, Arizona; and Chesterfield Inlet, Canada. A number of large electrical eddies appear here. The curves which outline them are so drawn that the same (but undetermined) amount of current flows between adjacent curves. *This does not apply to the central curves* or the dotted ones in *some of the weaker* eddies. Arrows show the direction in which the current flows. This system is fixed *with respect to the sun*, which is *directly* above *the center. As the earth rotates*, different aspects of the eddies appear at a given place. Take as an example Tucson, which is at about 32°N latitude and about 111°W longitude; *the local time is between 10 and 11* A.M.; one of the more intense tubes is over the place and the direction of flow is nearly due south. As this system of currents moves westward relative to the earth there will be little change at Tucson either in the intensity or the direction of the current until *about 13 h (1* P.M.*)* when the current begins to weaken and to veer towards the west; at about 16 h it is weaker and the *direction is NNE. One can* follow the diurnal variation at Watheroo (near the western tip of Australia) *in the same way. This* chart is not accurate for a number of other places. It would better represent the observations at other places if the tubes of flow were regarded as very flexible, easily deformable, so that the shape of the tubes might be changed and even the centers displaced all in such a way as to conform with the distribution of the electrical properties of the earth, especially in the more immediate environment of a given eddy. These eddies change from

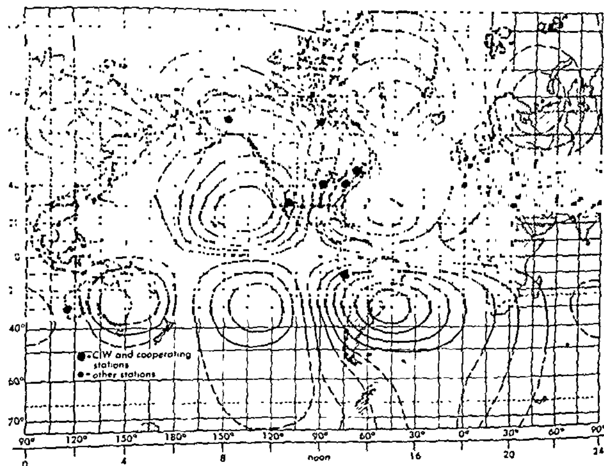


Fig 4. World view of a system of electric currents in the earth which would produce the annual average solar diurnal variation in earth-current potential-gradient at Watheroo, Western Australia; Tucson, Arizona; and Chesterfield Inlet, Canada, on quiet days.

a g
th

ted to an ...
aphic rock.) Any of these forms which may
t conductors of electricity. Cold dry ice, for
e, has a resistivity of 10^9 ohm-cm for direct
and low frequency alternating current; and
ohm-cm at 60 kilocycles. The dielectric
t for low frequencies at 0°C is 74.6 and for
requencies and very low temperatures it
o 3.0. Ice is also a poor conductor of heat
of the thermal constants of various forms of
terrestrial water under normal conditions is
le 2 (as compiled from various sources for
nean Icefield Research Program by M. M.
1954). The general conductivities for snow
erage firn noted in this table are based on
from the U.S. Army Corps of Engineers
, Ice and Permafrost Establishment, 1951).
ensity of firn ice is chosen arbitrarily between
values for average firn and ice. For compari-
he thermal properties listed in the Interna-
Critical Tables for rubber, steel, aluminum,
opper are also noted. The diffusivity figures
s table are rounded off for convenient refer-

Glaciers; Ground water, Hydrology, Sur-
water. [M.M.M.]

Bibliography: N. E. Dorsey, *Properties of Ordi-
Water-Substance*, ACS Monograph 81, re-
1954; R. F. Flint, *Glacial and Pleistocene
gy*, 1957; J. W. Glen, The creep of polycrys-
ice, *Proc. Roy. Soc. A*, 228:528–529, 1955;
M. Miller, *Glaciothermal Studies on the Taku
er, Alaska*, Assoc. Int. Hydrol. Publ. 39, 1954;
ew of the Properties of Snow and Ice, Snow,
and Permafrost Research Establishment, U.S.
y Corps of Engineers Rept. 4, 1951

restrial gravitation

attraction of the earth's mass at or above a
t on the earth's surface. The point is commonly

...

avitation). This assumes that all the earth's
s is concentrated at its gravity center C. Then
gravitation at P may be said to equal the New-
ian gravitational constant k multiplied by the
s M divided by the square of the distance r
m the point P to the earth's center C. The math-
atical expression is $G = kM/r^2$, wherein $k =$

...

ably varying determinations by more than a
zen different scientists. A more accurate value
f the earth's gravitation G may be gained by ex-
nding the integral

$$\Gamma = k \int_M dm/e^2$$

over the earth; for this, e is the distance of the mass element dm from the computation point P. The gravitation is used in celestial mechanics, for instance, in the computation of the moon's orbit.

Gravity. In the gravimetric and geodetic studies of the earth, consideration of the force due to gravitation alone is not sufficient. It is necessary to consider also the effect of the centrifugal force, $c = \omega^2 r \sin \vartheta$, of the earth's diurnal rotation, where ω is the angular velocity of the earth, r its radius, and ϑ the geocentric colatitude of the computation point. Gravity g is the difference: gravitation minus effect of the centrifugal force, or

$$g = k \int_M dm/e^2 - \omega^2 r \sin \vartheta$$

In fact, g is the acceleration of the earth's gravity, but it is commonly and simply called gravity and is understood to have the dimension of the acceleration.

MEASUREMENTS AND OBSERVATIONS

Gravity measurements. The acceleration of gravity, briefly termed gravity or g, can be measured either by dynamic or static apparatuses. The most used dynamic method measures the period of a swinging body under the attraction of the gravity force, by means of the pendulum, inverted pendulum, or the gravity variometer. Recently also, the measurement of the velocity of a falling body has been used. The static group consists of the spring balance principle (the usual gravimeters) and the gas-pressure gravimeters. Gravity measurements are either absolute or relative.

Pendulum method. In gravity measurements, pendulum observations have been most used. The invar pendulums of nickel-steel alloy have replaced the older brass pendulums because the length l of the invar pendulum is almost independent of the temperature variations. On the other hand, the crystalline structure of the invar, like that of the alloys in general, is ... in pendulum l... the magnetic effe... be used to protect the pendulum against such effects.

The single pendulum apparatus is the simplest in principle. It can hardly be used, however, because the sway of the fundamental earth base cannot be easily eliminated, and the changes of pendulum length cannot be estimated at all. Therefore, the dual- and multiple-pendulum apparatuses are applied in practice. When two pendulums swing in the same plane and have almost the same length, the effect of the sway can be eliminated. Changes of length can also be discovered except in the most unusual circumstance that both pendulums are changed the same amount and in the same direction. In four-pendulum apparatuses, both couples of the pendulums swing in different planes, 90° from one another. The accuracy of this method is 1 milligal (mgal), and of the best apparatuses of Cambridge and the Gulf Oil Co., it may even be 0.3–0.5 mgal (1 gal is equivalent to an acceleration of

the world's glaciers in mountains and subpolar regions. This is to be compared with approximately 32% of the world's land covered by glaciers during their maximum extension in the Pleistocene Ice Age. A close estimate of the total volume of terrestrial ice is not possible because the thickness of the Antarctic ice sheet is not yet adequately known. Flint, in 1957, tentatively estimated the volume of glacier ice existing today as about 24,000,000 km³, "equivalent to a layer of water having the area of the present oceans and approximately 59 m thick." No estimate has been attempted of the volume of subsurface terrestrial frozen water existing in permafrost regions.

Properties of terrestrial ice. Ice is one of the more abundant minerals on the surface of the earth. It is usually observed as colorless and transparent, but in large, dense masses it shows a vivid light blue color. The other physical properties of terrestrial frozen water vary considerably under changed conditions of internal temperature, load, crystalline orientation, and mass density. For example, ice has an indentation hardness which varies with the mass temperature and a scratch microhardness which differs with the orientation of the crystal plane. On the microhardness or Mohs scale at 0°C it has a hardness of about 2 and at −44°C a hardness of about 4 (*see* HARDNESS SCALES). Coincident with this property is its variable plasticity or ability to deform under stress. The deformation rate is considerably reduced under colder temperatures. At a temperature of −1°C, the deformation (flow) rate of polycrystalline ice may b[e] expected to be slowed to approximately one-fifth o[f] that at 0°C. For example, ice chilled to −10°C ha[s] been observed by J. W. Glen to have 25 times a[s] much ability to resist deformation as ice at 0°C.

Special hydrothermal relationships. No significant influence on the deformation rate appears in ice through changes in hydrostatic pressure, the relationship being as negligible as in liquids; but the melting point decreases as the pressure increases. This amounts to a reduction in temperature of 0.0075°C for each atmosphere of increase in pressure. As a result of this property, the melting temperature of an ice mass at any depth is fundamentally conditioned by the weight of the overlying mass. The term used to refer to this is pressure melting temperature. It is because of this characteristic that ice skating is possible. At the edge of a skate blade sufficient pressure is exerted to cause formation of a thin film of water, which serves as a lubricant. Similarly, when a wire is pressed into a block of ice, or even when pieces of ice are pressed together, melting occurs at the contact surface. When the pressure is released, the water refreezes, uniting the ice into a contiguous mass. This is the process of regelation.

Classification of ice forms. Although there is only one ordinary form or phase of terrestrial frozen water, ice-I, five other stable phases of ice exist, in addition to one unstable phase. These, however, are only the product of great pressure and are not found in normal conditions outside of the experimental laboratory. The stable, extraordinary forms are known as ice-II, III, V, and VII. Ice-IV is the unstable category. Each of these reverts to normal terrestrial ice, water, or both, when the pressure is released. Table 1 shows pressure and temperature parameters (triple points) under which these forms of ice and related liquid water exist.

Massive ice constants. Because massive ice is an aggregate of individual ice crystals, it may also be considered a monocrystalline rock. In the form of snow or firn (granular summer equivalent), it

Table 1. Temperature and pressure relationships for forms of ice and related liquid water*

Forms	Temperature, °C	Pressure, atm
Water, ice-I, III	−22.0	2,047
Ice-I, II, III	−34.7	2,100
Water, ice-III, V	−17.0	3,417
Ice-III, II, V	−24.3	3,397
Water; ice-V, VI	+ 0.16	6,175
Water, ice-VI, VII	+81.6	21,700

* Based on data reported in 1940 by N. E. Dorsey

Table 2. Thermal constants of various forms of frozen water compared with liquid water and other substances*

Water forms and other substances	Conductivity, cal/(°C)(cm)(sec)	Specific heat, cal/(°C)(g)	Density, g/cm³	Thermal diffusivity, cm²/sec	Relative diffusivity to ice (approx ratio)
Frozen water					
New snow	0.0003	0.5	0.20	0.0030	0.27
Old snow	0.0006	0.5	0.30	0.0040	0.36
Average firn	0.0019	0.5	0.55	0.0070	0.64
Firn ice	0.0038	0.5	0.75	0.0100	0.91
Ice	0.0050	0.5	0.92	0.0110	1
Comparative materials					
Water (0°C)	0.0014	1.0	1.00	0.0014	0.13
Rubber	0.0005	0.40	0.92	0.0014	0.13
Steel (mild)	0.1100	0.12	7.85	0.12	11
Aluminum	0.4800	0.21	2.70	0.86	78
Copper	0.9300	0.09	8.94	1.14	104

* Based on data from U.S. Army Corps of Engineers, *Review of the Properties of Snow and Ice*, 1951, and other sources as reported in M. M. Miller, *Glaciothermal Studies on the Taku Glacier, Alaska*, 1954.

has been genetically likened to a sedimentary rock. (Similarly pond, river, sea, or refrigerator ice may be likened to an igneous rock, and glacier ice to a metamorphic rock.) Any of these forms when dry are poor conductors of electricity. Cold dry ice, for example, has a resistivity of 10^9 ohm-cm for direct current and low-frequency alternating current; and 3×10^6 ohm-cm at 60 kilocycles. The dielectric constant for low frequencies at 0°C is 74.6 and for high frequencies and very low temperatures it drops to 3.0. Ice is also a poor conductor of heat. A list of the thermal constants of various forms of frozen terrestrial water under normal conditions is in Table 2 (as compiled from various sources for the Juneau Icefield Research Program by M. M. Miller, 1954). The general conductivities for snow and average firn noted in this table are based on data from the U.S. Army Corps of Engineers (Snow, Ice and Permafrost Establishment, 1951). The density of firn ice is chosen arbitrarily between given values for average firn and ice. For comparison, the thermal properties listed in the International Critical Tables for rubber, steel, aluminum, and copper are also noted. The diffusivity figures in this table are rounded off for convenient reference.

See GLACIER; GROUND WATER; HYDROLOGY, SURFACE WATER. [M.M.M.]

Bibliography: N. E. Dorsey, *Properties of Ordinary Water Substance*, ACS Monograph 81, reprint 1954; R. F. Flint, *Glacial and Pleistocene Geology*, 1957; J. W. Glen, The creep of polycrystalline ice, *Proc. Roy. Soc. A*, 228:528–529, 1955; M. M. Miller, *Glaciothermal Studies on the Taku*

Army Corps of Engineers Rept. 4, 1951.

Terrestrial gravitation

The attraction of the earth's mass at or above a point on the earth's surface. The point is commonly designated P and the terrestrial gravitation is abbreviated to C. The major component of this gravitation is specified according to Newton's law (see GRAVITATION). This assumes that all the earth's mass is concentrated at its gravity center C. Then the gravitation at P may be said to equal the Newtonian gravitational constant k multiplied by the mass M divided by the square of the distance r from the point P to the earth's center C. The mathematical expression is $G = kM/r^2$, wherein $k = 6.673 \times 10^{-8}$ cm^3 g^{-1} sec^{-2} (cgs units). The value of 6.673 (1942 by P. R. Heyl and P. Chrzanowski) is selected as possibly most reliable from numerous slightly varying determinations by more than a dozen different scientists. A more accurate value for the earth's gravitation G may be gained by expanding the integral

$$\int dm/e^2$$

over the earth; for this, e is the distance of the mass element dm from the computation point P. The gravitation is used in celestial mechanics, for i... in the computation of the moon's orbit.

...sider also the effect of the centrifugal force, $c = \omega^2 r \sin \vartheta$, of the earth's diurnal rotation, where ω is the angular velocity of the earth, r its radius, and ϑ the geocentric colatitude of the computation point. Gravity g is the difference: gravitation minus effect of the centrifugal force, or

$$g = k \int_M dm/e^2 - \omega^2 r \sin \vartheta$$

In fact, g is the acceleration of the earth's gravity, but it is commonly and simply called gravity and is understood to have the dimension of the acceleration.

MEASUREMENTS AND OBSERVATIONS

Gravity measurements. The acceleration of gravity, briefly termed gravity or g, can be measured either by dynamic or static apparatuses. The most used dynamic method measures the period of a swinging body under the attraction of the gravity force, by means of the pendulum, inverted pendulum, or the gravity variometer. Recently also, the measurement of the velocity of a falling body has been used. The static group consists of the spring balance principle (the usual gravimeters) and the gas-pressure gravimeters. Gravity measurements are either absolute or relative.

Pendulum method. In gravity measurements, pendulum observations have been most used. The invar pendulums of nickel-steel alloy have replaced the older brass pendulums because the length l of the invar pendulum is almost independent of the temperature variations. On the other hand, the crystalline structure of the invar, like that of the alloys in general, is variable and can cause small changes in pendulum lengths. In addition, it is sensitive to the magnetic effect, so that a Helmholtz coil must be used to protect the pendulum against such effects.

The single pendulum apparatus is the simplest in principle. It can hardly be used, however, because the sway of the fundamental earth base cannot be easily eliminated, and the changes of pendulum length cannot be estimated at all. Therefore, the dual- and multiple-pendulum apparatuses are applied in practice. When two pendulums swing in the same plane and have almost the same length, the effect of the sway can be eliminated. Changes of length can also be discovered except in the most unusual circumstance that both pendulums are changed the same amount and in the same direc-

...ligal (mgal), and of the best apparatuses of Cambridge and the Gulf Oil Co., it may even be 0.3–0.5 mgal (1 gal is equivalent to an acceleration of

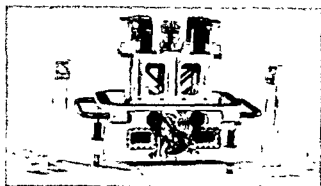


Fig. 1. Gulf minimum-pendulum apparatus An extra pair of fused-quartz pendulums and pyrex knife-edge flats illustrate the design of the minimum pendulums. (Gulf Research and Development Co.)

1 cm/sec², and a milligal is the equivalent of 0 001 cm/sec²). The most used pendulums are the ½-sec pendulums of the Sterneck type (Fig 2) because they are relatively short, with a length of 0 25 m These pendulums swing in a high vacuum to minimize air-induced errors.

The differential equation of the pendulum method is $\ddot{\varphi} + (g/l)\varphi = 0$, where φ is the elongation angle of the pendulum, $\ddot{\varphi}$ its second derivative with respect to time, l its length, and g the gravity From this equation is obtained the main term for the observed swinging time T

$$T = \pi \sqrt{\frac{l}{g}}$$

Since the elongation angle of the pendulum is not infinitely small, but has a value φ (about $1' - 2'$), the following arc correction ΔT is needed to get the corrected value T_0·

$$T_0 = T - \frac{\varphi_0^2}{16} T$$

Gravity observations at sea. If the suspension point of the pendulum also has a horizontal acceleration $\ddot{\gamma}$, then the following equations of two forced pendulums of the length l_1 and l_2, swinging in the same apparatus and in the same plane but in different phases, can be derived

$$\begin{cases} \ddot{\varphi}_1 + \dfrac{g}{l_1}\varphi_1 + \dfrac{1}{l_1}\ddot{\gamma} = 0 \\ \ddot{\varphi}_2 + \dfrac{g}{l_2}\varphi_2 + \dfrac{1}{l_2}\ddot{\gamma} = 0 \end{cases}$$

If $l_1 = l_2 = l$, it follows by subtraction that the equation is

$$(\varphi_2 - \varphi_1)'' + \frac{g}{l}(\varphi_2 - \varphi_1) = 0$$

This describes a fictitious free pendulum of the elongation angle $(\varphi_2 - \varphi_1)$ and of the length l of the original pendulums.

This equation is the principle of the Vening Meinesz sea pendulum apparatus, which can eliminate the effect of the horizontal acceleration $\ddot{\gamma}$ of the boat. Because the length of two pendulums will in fact not be exactly the same, a small correction is needed. To safeguard the accuracy of results, three pendulun should be used, designated 1, 2, and 3, and allowe to swing in the same plane as that of two fictitiou pendulums with the elongation angles $(\varphi_3 - \varphi_1)$ an $(\varphi_3 - \varphi_2)$; two values for g will be obtained. By ½ hour observations made in a submarine at a dept

1000 points
lumbia Univ
L. Worzel c
tions at sea
4000 points.
made possible the geodetic applications of the gravimetric method.

Sea and underwater gravimeters. During the late 1950s, the test measurements of L. Worzel, using the A. Graf sea gravimeter, and the experiments made in the Gulf of Mexico with the L. LaCoste sea gravimeter from a surface boat have shown that these apparatuses yield almost the same accuracy as submarine measurements.

The Gulf Oil Company's underwater gravimeter is used in waters to about 200-m depth (Fig. 3). Most of the apparatus is at the bottom of the sea during the measurement, but the results are registered in a surface boat. Measurement methods are also being developed in Italy, the Netherlands, and Finland to measure gravimetrically the shallow waters of the Adriatic, the North, the Barents, and the Baltic seas. These methods were very successful, yielding nearly the accuracy of gravimeter observations on land. T. Honkasalo has been able to measure gravity at 178 points in the Baltic Sea area and at 24 points in the Barents Sea, even during a storm of 6 Beaufort In using these gravimeters, no

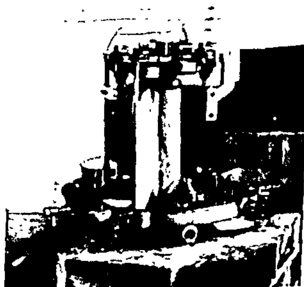
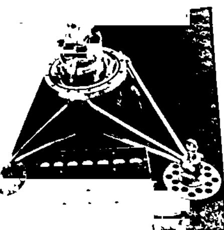


Fig. 2. Sterneck's four-pendulum apparatus. One pendulum of each pendulum pair is shown. In operation, the apparatus is in a high-vacuum flask. Two pendulums swing in the same plane to eliminate the sway of the stand.



. 3. The Gulf underwater gravimeter and remote trol. Watertight housing cover removed. (Gulf Research and Development Co.)

ubmarines are needed, and necessary computation work is comparatively light.

In all observations at sea—except in the underwater gravimeter measurements—the Eotvos effect, or the effect of the east-west component of the vessel's velocity, results in an error of 4 cos φ mgal for the velocity error of 1 knot per hour; an error of this magnitude could easily be caused by the ocean currents. Therefore, all methods which can yield an accuracy of some few mgal are good for gravity observations at sea.

Absolute gravity measurements. In these meas-

Both l and T are needed for computation of g

$$T^2 = \pi^2 l/g$$

Falling body methods. Since it is now possible to measure the time intervals with very high accuracy (even 10^{-11} sec), the falling body method, introduced by Galileo, is used in more modern form In this method, a bar of 1- or 2-m length is allowed to fall in a vacuum, and the times are recorded when different marks on the bar pass the observation level. This method has been used by C. Volet in Paris since 1952, by E. Rieckman in Braunschweig; it has been used since 1955 in Ottawa and in Leningrad since 1956. The accuracy is of the order 1-5 mgal In Ottawa, an accuracy better than 1 mgal is claimed.

Other absolute methods. Bessel's method of the thread pendulum is also used. New absolute gravity measurements have also been made in Teddington, Potsdam, Helsinki, Washington, and Madison. Wisconsin.

Relative gravity observations. Since the absolute measurements are difficult and time-consuming, relative methods are most used. In these measurements, the ratio between the gravity g_0, measured at the gravity base station, and g, measured at the gravity field station, is obtained. The equation of the relative measurements is the identity: $g = g_0 g/g_0$. If g_0 is known, the ratio g/g_0 obtained from the observations gives the gravity g.

In the relative observations, the length l, which is difficult to measure, disappears, and only the swinging times T and T_0 are needed, as can be seen from the formulas

$$g = T^2 l/\pi^2 \qquad g_0 = T_0^2 l/\pi^2 \qquad g/g_0 = T_0^2/T^2$$

Gravimeters. Earlier relative gravity observations were done mostly by the pendulum method already described, but now gravimeters are used almost exclusively.

The gravimeters with a spring balance system have replaced the Haalck gas pressure gravimeter and the Holweck inverted pendulum method used by P. Lejay in his global gravimetric explorations in France, North Africa, and the Middle and Far East.

The usual type of gravimeter is based either on the spring balance principle (stable system) or astatic balance principle (unstable system)

The principle of the stable system is simple. When a weight hangs from a spring, the length of the spring will change with variation of the gravity g The greater the gravity, the longer the spring. If the other conditions, particularly the elastic properties of the spring, remain unchanged, relative values of g can be obtained by measuring the spring length at the observation point The Hartley gravimeter, the Hoyt Gulf gravimeter, and the Nørgaard gravimeter are the best known representatives of the stable type.

The principle of the unstable gravimeters is more difficult to understand The force of gravity is kept

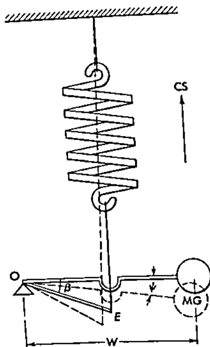


Fig 4. Principle of the unstable gravimeter. A mass MG at the end of a beam with length W is pivoted at O The moment of its weight $MGW \sin\psi$ is balanced by the balancing couple, also a function of ψ.

in unstable equilibrium with the restoring forces by a third force which enlarges the effect of any change of gravity from its equilibrium value (Fig. 4). The moving system has an arm hinged at one end and carries the weight MG at the other. In contrast to the stable gravimeters, the spring is not attached to the weight arm itself, but to a lever arm E at an angle β with respect to the weight arm W. If the angle β is large, the spring-restoring-torque curve becomes nonlinear. The instrument works close to the instability point so that its sensitivity is very high. The gravimeters of LaCoste-Romberg, Frost, Magnolia, North American, and Worden are of the unstable type.

When gravimeters are used for exploration purposes to study the gravity differences of relatively small areas, the range of the gravimeter need not be large. This is in contrast to the geodetic gravimeters, the range of which must be large because they must measure gravity differences of many hundreds and even several thousands of milligals. The Nørgaard gravimeter, range 2000 mgal, and the Worden gravimeter, range 5000 mgal, are the best among the geodetic gravimeters. The Worden can measure as much as the difference between the gravity at the pole and at the equator. Figure 5 shows the Worden gravimeter. Its diamete[r] cm, height 29 cm, weight 2.3 kg and, with th[e car]rying case and base plate, 5.4 kg. The W[orden] gravimeter has been mostly used in the worl[d] gravity correlation trips of G. P. Woollard a[nd] group.



Fig. 5. Exterior view of Worden gravimeter. Diameter 12 cm, height 29 cm, weight 2.3 kg, with the carrying case and base plate 5.4 kg.

Gravity theory and formulas. To develo[p] theory of gravity, the rotating earth is assum[ed to] be a spheroid of equilibrium, and its figure, [tech]nically termed the geoid, is considered the eq[uipo]tential surface of the earth's gravity which [coin]cides with mean sea level. As both parts o[f the] gravity field (Newtonian attraction and varying portions of centrifugal force from rotation) a[ct in] radial direction to a point and in amounts of f[orce] dependent on the distance to the point, they h[illegible]

[illegible] ...rivation, now largely developed in more techni[cal] literature and summarized in certain mathemati[cal] derivations and summarizing formulas. Two of [the] most useful such summaries are Clairaut's form[ula] for polar flattening, with what is termed accura[cy] of the first order,

$$\alpha = \tfrac{5}{2}m - \beta \qquad (1)$$

in which m is the ratio of centrifugal force [to] gravity at the equator, γ_E, and β is the coefficien[t] of the principal latitude term; and the formula fo[r] theoretical gravity, with accuracy of second order

$$\gamma = \gamma_E(1 + \beta \sin^2 \varphi + \epsilon \sin^2 2\varphi) \qquad (2)$$

wherein φ is the latitude of the place or station and ϵ is the theoretical latitude constant 0.0000059.

Several gravity formulas have been derived on the basis of the gravity observations distributed on different latitudes around the world and reduced to sea level either by the free air or isostatic reductions. The equatorial gravity value γ_E and the coefficient β of the $\sin^2 \varphi$ term are derived from gravity anomalies, and ϵ is computed theoretically so that the *niveau* (leveled) spheroid applied coincides with the used reference ellipsoid. The table gives the parameters of some gravity formulas and the corresponding flattening α. From the parameters: $\gamma_E = 978.0490$ cm/sec² was derived by W. A. Heiskanen in 1928, $\beta = 0.0052884$ is computed from the flattening $\alpha = 1{:}297.0$ of the international ellipsoid, and $\epsilon = -0.0000059$ is obtained theoretically so that the niveau spheroid coincides with the international ellipsoid. From them is obtained the international gravity formula accepted in the Stockholm meeting of the International Union of Geodesy and Geophysics (IUGG) in 1930, and used in most countries, as follows:

$$\gamma = 978.0490\,(1 + 0.0052884 \sin^2 \varphi - 0.0000059 \sin^2 2\varphi)\ \text{cm/sec}^2$$

The geodetic, gravimetric, and satellite studies of recent years have indicated that the equatorial value γ_E of the International gravity formula has to be corrected by about −5 to −7 milligal (mgal, or 0.001 cm sec⁻²). The α value also has to be cor-

Parameters of some gravity formulas and corresponding flattening of meridian

Author	Year	Parameters			Flattening of meridian, α
		γ_E, cm/sec²	$\beta \times 10^3$	$\epsilon \times 10^6$	
Helmert	1901	978 030	5.302	−7.0	1:298 2
Bowie	1917	039	5 294	−7 0	1 297.5
Heiskanen	1928	.049	5.289	−7 0	1:297.06
Heiskanen	1938	0451	5 3026	−5.9	1 298 2
International	1930	.0490	5 2884	−5.9	1 297.0
Heiskanen, Uotila	1957	0496	5.2934	−5.9	1:297.4
Kaula	1958	.0406	5.3014	−5 9	1:298 15

by about −130 m, the α value corrected by −0 3%, and the β term increased by about

e following are five international quantities heir recommended, approximate, new values·

...ity and old value	Corrected value
...ional radius = 6,378,388 m	6,378,370 m
...ening of meridian 1 = 297 0	1 298 1
...ional gravity g = 978.0490 cm/sec²	978,045
...ficient of the sin² φ term = 0 0052884	0.005305
...ficient of the sin² 2φ term = −0 0000059	−0 0000059

...on of gravity with latitude is considered in grav... formulas, but gravity also differs along the same parallel, depending on whether the gravity has been measured in lowland, in high mountains, or at sea. Figure 6 indicates that the observed gravities g_0 and ... are not comparable with one another. Gravity g_1 is obviously smaller than g_0 because it is measured farther from the gravity center than g_0. Also, the ...

...e obtained by reducing them to the same surface. It is best, of course, to reduce to sea level or ge...

...ts only the effect of the elevation h of the observation point. The positive correction is $2gh/r$, in ...

...ection, by way of example, is for h = 1000 m, +308.6 mgal; for h = 3000 m, +925.8 mgal.

Bouguer reduction The mass of the mountain between the observation point for g_1 and sea level should also be considered. Here, the increase in gravity indicates that something has to be subtracted from the observed value. This is the Bouguer reduction which can be obtained from the formula

$$-\frac{3}{4}\frac{\rho}{\rho_m}\frac{g}{r}h$$

where ρ is the rock density and ρ_m the mean density of the earth. Using the values ρ = 2.67 and ρ_m = 5 52, the factor of the Bouguer reduction is $-0.1118h$. At stations within rough topography, it is also desirable to add the terrain correction, which is positive for the mountain top as well as in the valley as shown in Fig. 7. The sum of the free-air reduction and the Bouguer correction gives the Bouguer values for g: $g_B = (g + 0.3086h - 0.1118h + \text{terr. correction})$.

It is necessary also to remember to add to the oceans a mass of density 1.643 = 2.67 − 1.027. Thus, the density of the ocean (1 027 is the density of sea waters) would be changed to the same, 2 67,

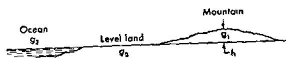


Fig. 6. Since the gravity values g_1, g_0, and g_2 are measured in different conditions, they cannot be compared with each other until they have been reduced to sea level

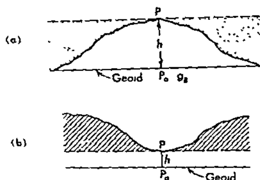


Fig. 7 Terrain correction is positive, both at the mountain top station and at valley station, because in (a), the first case, the Bouguer plate subtracts the nonexistent mass of the dotted part, in (b), the second case, the mass hatched area is above the gravity station and diminishes the gravity.

in unstable equilibrium with the restoring forces by a third force which enlarges the effect of any change of gravity from its equilibrium value (Fig. 4). The moving system has an arm hinged at one end and carries the weight MG at the other. In contrast to the stable gravimeters, the spring is not attached to the weight arm itself, but to a lever arm E at an angle β with respect to the weight arm W. If the angle β is large, the spring-restoring-torque curve becomes nonlinear. The instrument works close to the instability point so that its sensitivity is very high. The gravimeters of LaCoste-Romberg, Frost, Magnolia, North American, and Worden are of the unstable type.

When gravimeters are used for exploration purposes to study the gravity differences of relatively small areas, the range of the gravimeter need not be large. This is in contrast to the geodetic gravimeters, the range of which must be large because they must measure gravity differences of many hundreds and even several thousands of milligals The Norgaard gravimeter, range 2000 mgal, and the Worden gravimeter, range 5000 mgal, are the best among the geodetic gravimeters. The Worden can measure as much as the difference between the gravity at the pole and at the equator Figure 5 shows the Worden gravimeter. Its diameter is 1 cm, height 29 cm, weight 2.3 kg and, with the ca rying case and base plate, 5.4 kg. The Worde gravimeter has been mostly used in the world wid gravity correlation trips of G. P. Woollard and hi group.

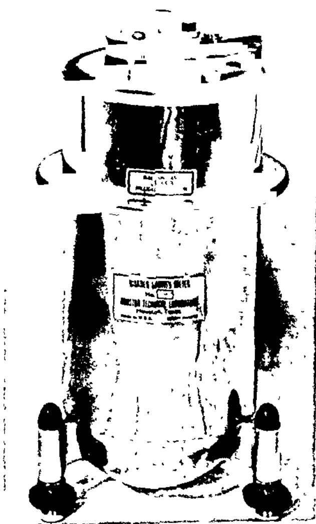


Fig. 5. Exterior view of Worden gravimeter. Diameter 12 cm, height 29 cm, weight 2 3 kg, with the carrying case and base plate 5.4 kg.

Gravity theory and formulas. To develop th theory of gravity, the rotating earth is assumed t be a spheroid of equilibrium, and its figure, tech nically termed the geoid, is considered the equipo tential surface of the earth's gravity which coir cides with mean sea level. As both parts of th gravity field (Newtonian attraction and varying pro portions of centrifugal force from rotation) act i radial direction to a point and in amounts of forc dependent on the distance to the point, they have potential which determines the shape of the geoid The potential may be considered in varying degree of complexity as to components and refinements o derivation, now largely developed in more technica literature and summarized in certain mathematica derivations and summarizing formulas. Two of th most useful such summaries are Clairaut's formul for polar flattening, with what is termed accurac of the first order,

$$\alpha = \tfrac{5}{2}m - \beta \qquad (1)$$

in which m is the ratio of centrifugal force t gravity at the equator, γ_E, and β is the coefficien of the principal latitude term; and the formula fo theoretical gravity, with accuracy of second order

$$\gamma = \gamma_E(1 + \beta \sin^2 \varphi + \epsilon \sin^2 2\varphi) \qquad (2)$$

wherein φ is the latitude of the place or statior and ϵ is the theoretical latitude constant 0.0000059.

Several gravity formulas have been derived on the basis of the gravity observations distributed on different latitudes around the world and reduced to sea level either by the free air or isostatic reductions The equatorial gravity value γ_E and the coefficient β of the $\sin^2 \varphi$ term are derived from gravity anomalies, and ϵ is computed theoretically so that the *niveau* (leveled) spheroid applied coincides with the used reference ellipsoid. The table gives the parameters of some gravity formulas and the corresponding flattening α. From the parameters: $\gamma_E = 978.0490$ cm/sec² was derived by W A Heiskanen in 1928, $\beta = 0.0052884$ is computed from the flattening $\alpha = 1{:}297.0$ of the international ellipsoid, and $\epsilon = -0.0000059$ is obtained theoretically so that the niveau spheroid coincides with the international ellipsoid. From them is obtained the international gravity formula accepted in the Stockholm meeting of the International Union of Geodesy and Geophysics (IUGG) in 1930, and used in most countries, as follows:

$$\gamma = 978.0490\,(1 + 0.0052884 \sin^2 \varphi - 0.0000059 \sin^2 2\varphi)\ \text{cm/sec}^2$$

The geodetic, gravimetric, and satellite studies of recent years have indicated that the equatorial value γ_E of the International gravity formula has to be corrected by about −5 to −7 milligal (mgal, or 0 001 cm/sec²). The α value also has to be cor-

Parameters of some gravity formulas and corresponding flattening of meridian

Author	Year	Parameters			Flattening of meridian, α
		γ_E, cm/sec^2	$\beta \times 10^3$	$\epsilon \times 10^6$	
Helmert	1901	978 030	5.302	−7.0	1:298 2
Bowie	1917	.039	5.294	−7.0	1 297.5
Heiskanen	1928	.049	5 289	−7.0	1:297.06
Heiskanen	1938	.0451	5.3026	−5 9	1:298 2
International	1930	.0490	5 2884	−5.9	1:297.0
Heiskanen, Uotila	1957	.0496	5 2934	−5.9	1:297.4
Kaula	1958	.0406	5 3014	−5.9	1:298.15

rected by about −130 m, the α value corrected by about −0.3%, and the β term increased by about 0.3%.

The following are five international quantities and their recommended, approximate, new values:

Quantity and old value	*Corrected value*
Equatorial radius $a = 6{,}378{,}388$ m	6,378,370 m
Flattening of meridian α 1: 297 0	1:298.1
Equatorial gravity $\gamma_E = 978.0490$ cm/sec^2	978,045
Coefficient of the $\sin^2 \varphi$ term $\beta = 0.0052884$	0.005305
Coefficient of the $\sin^2 2\varphi$ term $\epsilon = -0.0000059$	−0.0000059

... formulas, but gravity also differs along the same parallel, depending on whether the gravity has been measured in lowland, in high mountains, or at sea. Figure 6 indicates that the observed gravities g_0 and g_1 are not comparable with one another. Gravity g_1 is obviously smaller than g_0 because it is measured farther from the gravity center than g_0. Also, the gravity g_2 at sea is, or at least ought to be, smaller than g_0 because the density of the water is less than that of rock.

Gravity values comparable with one another may be obtained by reducing them to the same surface. It is best, of course, to reduce to sea level or geoid because this is an equipotential datum level. Such reduction can be carried out in different ways.

Free-air reduction. This type of reduction considers only the effect of the elevation h of the observation point. The positive correction is $2gh/r$, in which h is the elevation and r the radius of the earth. Its numerical value changes with the latitude, but the value $0.3086h$ is a good average and gives the correction in milligals. This considerable correction, by way of example, is for $h = 1000$ m, +308.6 mgal; for $h = 3000$ m, +925 8 mgal.

Bouguer reduction. The mass of the mountain between the observation point for g_1 and sea level should also be considered. Here, the increase in gravity indicat... ...ing has to be subtracted from the observed value. This is the Bouguer reduction which can be obtained from the formula

$$-\frac{3}{4}\frac{\rho}{\rho_m}\frac{g}{r}h$$

where ρ is the rock density and ρ_m the mean density of the earth. Using the values $\rho = 2.67$ and $\rho_m = 5.52$, the factor of the Bouguer reduction is $-0.1118h$. At stations within rough topography, it is also desirable to add the terrain correction, which is positive for the mountain top as well as in the valley as shown in Fig. 7. The sum of the free-air reduction and the Bouguer correction gives the Bouguer values for g: $g_B = (g + 0.3086h - 0.1118h + \text{terr. correction})$.

It is necessary also to remember to add to the oceans a mass of density $1.643 = 2.67 - 1.027$. Thus, the density of the ocean (1.027 is the density of sea waters) would be changed to the same, 2.67,

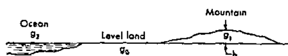


Fig. 6. Since the gravity values g_1, g_0, and g_2 are measured in different conditions, they cannot be compared with each other until they have been reduced to sea level.

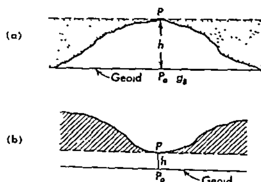


Fig. 7. Terrain correction is positive, both at the mountain top station and at valley station, because in (a), the first case, the Bouguer plate subtracts the nonexistent mass of the dotted part, in (b), the second case, the mass hatched area is above the gravity station and diminishes the gravity.

as the density of the rock. This is the Bouguer reduction at sea.

Isostatic reduction. In the isostatic reduction, it is assumed neither that the mountains are absolute mass surplus nor that the oceans present any absolute mass deficiency areas; hence, an isostatic equilibrium prevails. The mountain masses are considered compensated by the relatively light roots of the mountain, and the mass deficiency of the ocean basin is compensated by the heavy antiroots of the oceans, as diagrammed in Fig. 8. Such isostatic views offer compensations for the evident Bouguer gravity anomalies, which are known to be strongly negative in mountain areas and still more strongly positive at sea. The subterranean and submarine mass distribution of the earth's crustal rocks, therefore, is assumed to explain the surface anomalies by isostatic equilibrium and to yield a basis for computing adjustments or corrections for the surface gravity anomalies. After isostatic reduction, gravity values are those of a smoothed, fictitious earth, with neither mountains nor oceans.

The geoid of Europe. As an example of the gravimetrically computed geoids, the geoid of Europe, computed in the Mapping and Charting Research

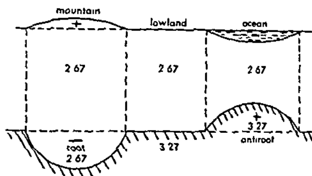


Fig. 8. Isostatically compensating light roots of the mountain and the heavy antiroots of the oceans. The under boundary of the earth's crust is an exaggerated mirror picture of the topography.

Laboratory of the Ohio State University in 1951, is diagramed in Fig. 9. The contour curves of 2 m interval are drafted on the basis of computations of N (distance between geoid and earth spheroid) at more than 1000 points, in corners of every square degree. This geoid covers the area between the northern latitudes 30° and 60° and between the longitudes 5°W and 30°E.

a

b

Fig. 9. Gravimetrically computed geoid of Europe. Contour interval: 2 m. The maps are obtained on basis of more than 1100 computation points.

Almost all N values of this geoid are positive. They increase gradually from the Gulf of Finland, where N is +10 m, to Spain and to the Mediterranean where it exceeds +40 m. The unusually lar . . . f the geoid contour between South Ita . . . di . . .

. . . cance for geodesy and also cast new light on . . . structure of the earth's crust

The absolute errors of the contour lines of this geoid might be about 5–10 m. This is brought about mostly by the effect of the gravimetrically unsurveyed areas of the Southern Hemisphere and of several large areas of the Northern Hemisphere. Only additional gravity observations can improve the results. The error of the shape of the geoid is much less, hardly more than one meter, because the effect of the unsurveyed areas far from Europe

. . . avity . . . t geodesy . . . also in geophysics. In fact, the gravity . . .

. . . particular area, only the gravity anomaly field of the area in question and of its neighborhood is specifically needed, but the gravity anomaly field of the whole earth must be known before the gravimetric method can be applied in physical geodesy

ISOSTASY

. . . as well as of the thickness of the earth's crust and its relationship to the topography

The topography of the earth with its mountains, valleys, ocean islands, and ocean basins indicates . . .

. . . the earth crust is in almost complete isostatic equilibrium By isostatic equilibrium is meant such distribution of the masses in the earth's interior that there is a certain depth (the depth of compensation) at which all surface units are under the same pressure regardless of whether the surface unit is under a mountain, lowland or ocean This is possible only when the mean density of the earth layers downward to the depth of compensation is smaller under the mountains and larger under the oceans than under the lowlands.

Main isostatic systems. Three isostatic systems have been developed the Pratt-Hayford system, the Airy Heiskanen system, and the Vening Meinesz system

Pratt Hayford system In this, it is assumed that the density of the earth's crust is smaller as the elevation increases. This is analogous to fermenting dough, the density of which will be smaller as the dough rises higher At the depth of compensation, the surface units are under the same pressure or, what is almost the same, have the same mass, regardless of where the mass unit is located. Figure 10 explains the idea of this system. The original Pratt's hypothesis is given in Fig 11.

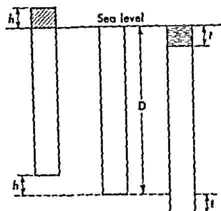


Fig. 10. Pratt-Hayford isostatic system. To simplify the isostatic computations of Pratt Hayford, Hayford computed the depth of compensations from the physical surface of the earth instead of from sea level

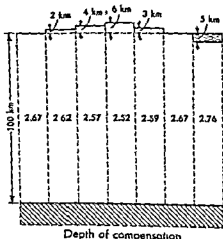


Fig 11. Idea of Pratt's original isostatic hypothesis Smaller density of the mountain columns and greater density of the ocean columns compensate the mass surplus of the mountains and mass deficiency of the oceans.

Airy-Heiskanen system. This system is in sharp contrast to the Pratt-Hayford system It postulates that the higher the mountains, the deeper they are sunk into the underlayer, and that under the oceans are antiroots of heavy material. Mountains are assumed to be floating in the heavier substratum as icebergs float in the oceans. The light roots of a mountain compensate the mass of the mountain. In a similar way, the heavy antiroots of the oceans compensate the mass deficiency of the ocean itself. The underboundary of the crust is an exaggerated reverse image or mirror picture of the topography. The higher the mountain, the deeper the roots; the deeper the ocean, the thicker the antiroot. Figure 12 illustrates this system. The original idea of

as the density of the rock. This is the Bouguer reduction at sea.

Isostatic reduction. In the isostatic reduction, it is assumed neither that the mountains are absolute mass surplus nor that the oceans present any absolute mass deficiency areas; hence, an isostatic equilibrium prevails. The mountain masses are considered compensated by the relatively light roots of the mountain, and the mass deficiency of the ocean basin is compensated by the heavy antiroots of the oceans, as diagrammed in Fig. 8. Such isostatic views offer compensations for the evident Bouguer gravity anomalies, which are known to be strongly negative in mountain areas and still more strongly positive at sea. The subterranean and submarine mass distribution of the earth's crustal rocks, therefore, is assumed to explain the surface anomalies by isostatic equilibrium and to yield a basis for computing adjustments or corrections for the surface gravity anomalies. After isostatic reduction, gravity values are those of a smoothed, fictitious earth, with neither mountains nor oceans.

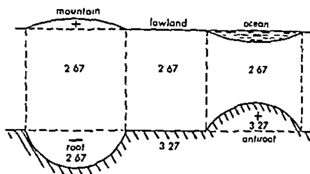


Fig. 8. Isostatically compensating light roots of the mountain and the heavy antiroots of the oceans The under boundary of the earth's crust is an exaggerated mirror picture of the topography

The geoid of Europe. As an example of the gravimetrically computed geoids, the geoid of Europe, computed in the Mapping and Charting Research Laboratory of the Ohio State University in 1951, is diagramed in Fig. 9. The contour curves of 2-m interval are drafted on the basis of computations of N (distance between geoid and earth spheroid) at more than 1000 points, in corners of every square degree. This geoid covers the area between the northern latitudes 30° and 60° and between the longitudes 5°W and 30°E.

a

b

Fig. 9. Gravimetrically computed geoid of Europe. Contour interval. 2 m. The maps are obtained on basis of more than 1100 computation points.

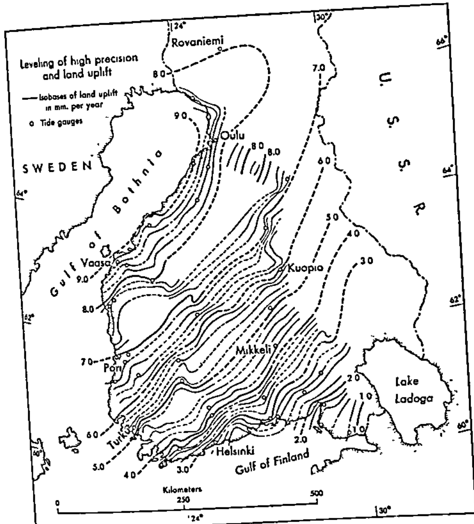


Fig 16. Land uplift in Finland according to E Kaariainen Figures 9.0, 8 0, 7.0, and so on give the isobases of land uplift in mm per year

prevails. Best evidence for this is the fact that the thickness of the ice layers is in many places much higher than the elevation of the glacier above the ocean level

In Fennoscandia, there is another phase of such a so-called experiment of nature. During the glacial period, the earth's crust was depressed about 700 m. At the end of the glacial period, the crust upraised about 250 m, after that, it is uplifted 270 m in the area of highest upheaval. Since the earth's crust is not yet in equilibrium, it is still rising about 90 cm per 100 years This uplift of land will continue until an equilibrium is almost reached.

Figure 16 shows the rate of the land uplift in Finland computed from the elevation differences obtained from two precise levelings at an interval of about 50 years. [W A.H.]

Bibliography: D. E. Gray (ed.), *American Institute of Physics Handbook*, 1957; W. A Heiskanen, ... Heiskanen, *The Earth and Its Gravity Field*, 1958, B F Howell, Jr., *Introduction to Geophysics*, 1959; G P Kuiper, *The Solar System*, vol. 2, 1954

Terrestrial magnetism

The natural magnetism of the earth. The designation geomagnetism is now given some preference over the older term, terrestrial magnetism. For a more detailed discussion, *see* GEOMAGNETISM; *see also* AEROMAGNETIC SURVEYING; COMPASS, MAGNETIC, DECLINOMETER; EARTH INDUCTOR; MAGNETISM, MAGNETOMETER

Terrestrial nuclear reactions

taneous reactions ... spontaneous fission of the heavier elements; and (2) induced reactions produced by the interaction of an incident particle (projectile) on a target nucleus. The naturally occurring flux of projectiles capable of producing induced nuclear reactions in the earth and its atmosphere consists of the alpha, beta, and gamma rays emitted by radioactive iso-

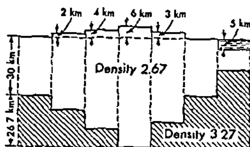


Fig. 12. Airy-Heiskanen isostatic system. The mountains float in the denser substratum as icebergs do in the ocean. The light roots of the continents and the heavy antiroots of the oceans *compensate* the topographic masses.

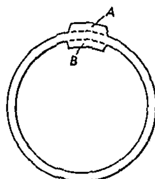


Fig. 13 Airy's original isostatic assumption. The root B compensates the mass of high plateau A.

Airy in 1855 is shown in Fig. 13. In this assumption, which is in agreement with seismic evidence, a constant compensating density of $\Delta\rho = 0.6$ is used to compute the thickness t of the root and the thickness t' of the antiroot. The equation expressing the compensation in the continents is $\rho \cdot h = \Delta\rho \cdot t$, or $t = \rho h/\Delta\rho$, where h is the elevation of the topography. In the ocean areas, the equations are as follows: $(\rho - 1.03)h' = \Delta\rho \cdot t'$, or $t' = (\rho - 1.03)\, h'/\Delta\rho$, where h' is the depth of the ocean; the quantity 1.03 is the density of the ocean water.

The obtained gravity anomalies Δg are expressed in the form $\Delta g = a + b \quad h$, where a and b are constants to be computed from the gravity anomalies. One assumes the normal thickness T of the earth's crust to be most plausible, which will give for b the value zero; this means that the obtained gravity anomalies are independent of the elevation of the topography. Using this criterion, Heiskanen and his students have obtained the following values for most plausible T values: in Norway 38 km, in Ferghana basin 38 km, in the East Alps 20 km, in the West Carpathians 30 km; C. Morelli has obtained for North Italy the value 29 km, Hales for South Africa 30 km. When to these values the average elevation of different mountains and the corresponding thickness of the root are added, the thicknesses of the whole earth crust under the mentioned areas are about 43 km, 42 km, 31 km, 33 km, 36 km, and 37 km.

The obtained values agree well with the Mohorovičić or *M* discontinuity of the earth's interior, computed by seismic methods.

The tendency of the earth's crust seems to be toward isostatic equilibrium. In many cases, the equilibrium is already reached; in some other interesting areas, there are such complicating factors as tectonic phenomena, loading and unloading of the glacial ice sheet, sediment layers, and others, which prevent the crust from reaching an equilibrium.

The general root theory. In the normal case, the mountains have normal roots, which means that the equilibrium prevails and gravity anomalies are al- ... for instance the ... slands, are too ... ese are examples of mountain or island with too small a root formation and unusual positive gravity anomalies. There are also areas, such as the belts of negative anomalies of the East Indies, the West Indies, and eastward off Japan, where there seem to be root formations but no mountains, and systematically negative gravity anomalies

Isostatic readjustments in nature. The postglacial land uplift in Fennoscandia and of the bowls of the glacier areas in Greenland and in Antarctica are examples of nature's current great isostatic "experiments." In Greenland and in Antarctica, the earth's crust has sunk under the iceload of 2–3 km thickness, apparently until isostatic equilibrium

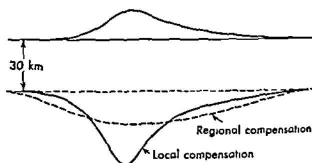


Fig 14 In Vening Meinesz's regional isostatic-floating theory, the compensating masses of the mountains (and of the oceans) are distributed broadly in the horizontal direction. The topography is a load on an unbroken crust or, in ocean areas, a load taken away from the crust. This illustrates the difference between the local and regional compensation

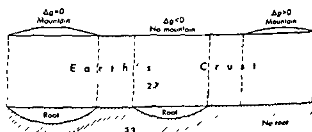


Fig. 15. General root theory. If the mountains have normal roots, isostatic equilibrium prevails. When there are root formations without any mountains, or mountains without any roots, there is no isostatic compensation and gravity anomalies Δg are negative and positive, respectively.

sion of space also, with birds of similar food requirements occupying different levels with independent patterns of territories and without direct competition for food. Possession of a territory as-

...

torial behavior also imputes some ...

rd population, and may have evolutionary advantage in preventing over-crowding From year to

...

... of one-year fertile males that are unmated but

...

area is defended, but the larger feeding area is not.

In many mammals and reptiles, and some other animals, individuals, pairs, or social groups move over a fairly definite area in seeking food or when engaged in other activities. Such an area is termed a home range but it is not termed a territory unless it is defended against others of the species See BEHAVIOR AND HEREDITY; SOCIAL ANIMALS.

[R.H.W.]

Bibliography: L. R. Dice, *Natural Communities*, 1952.

Tertiary

The earlier major division of Cenozoic time (Cenozoic Era), extending from the end of the Cretaceous (end of Mesozoic) to the Quaternary, or the later period of the Cenozoic. The tertiary is character-

PRE-CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBON-IFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	Mississippian	Pennsylvanian	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

ized particularly by (1) the rapid development of mammals and birds, grasses, higher types of flowering plants, and certain groups of invertebrates, (2) development of the modern configuration of the continental landmasses, and (3) widespread volcanic activity Because of the great numbers and large size of its mammals, the Tertiary period is often called the age of mammals. See CENOZOIC

The Tertiary System includes all rocks formed during the Tertiary period, but it is used most specifically with reference to the sedimentary rocks formed during this time, which contain the plant and animal remains that constitute the primary basis for identification.

Lithology. The Tertiary rocks include all common sedimentary types, both marine and continental as well as intermediate types, and are, in a gross manner, characterized by their unconsolidated or partly consolidated nature and by the frequent occurrence in them of large numbers of well-preserved fossils. The Tertiary rocks are widespread in many parts of the world. Igneous types are best known in mountainous areas; nonmarine strata are most widely distributed in the central parts of continental land areas, and marine sedimentary materials are most abundant around the continental margins near sea level, where they underlie large segments of the modern coastal plains and continental shelves. Marine Tertiary strata are also found at high elevations, for example in the Himalayas and the central plateau of Chiapas (Mexico), where they have been uplifted by various earth movements

Tertiary volcanic rocks are thousands of feet thick in some mountainous areas of the world Nonmarine sedimentary deposits are generally thin, although great thicknesses are present in some of the Rocky Mountain basins and in southeastern Europe The marine Tertiary is very thick in some places, being in the order of 20,000–25,000 ft in parts of the Gulf of Mexico coastal area and in certain sedimentary basins of western North America

Local names have been applied to the many rock units which compose the Tertiary System, and, as a result, students of the Tertiary are confronted by a bewildering array of formational names.

Limits and subdivisions. The boundaries of the Tertiary system are marked in some places by physical breaks in the sedimentary record. Elsewhere the sequence of strata contains no break and reflects continuous deposition from the Cretaceous period into the Tertiary, or from the Tertiary into the Quaternary In these instances, division of Tertiary strata from Cretaceous or Quaternary ones is based arbitrarily on faunal or other evidence, if available, or on the needs and requirements of geologic work in the particular area concerned.

The principal worldwide divisions of the Tertiary, established mainly through studies of these deposits in western Europe, are the Pliocene, Miocene, Oligocene, Eocene, and Paleocene (see separate articles by these titles)

Stratigraphic nomenclature. Data obtained by Arduino in Italy, by Cuvier, Brongniart, and Deshayes in France, by D'Halloy in Belgium, and by T Webster, W Buckland, and C Lyell in England, were used by Sir Charles Lyell in 1833 to divide the Tertiary into Eocene (oldest), Miocene, and Pliocene, based on the percentage of living species contained in the deposits Subsequently, the term Oligocene was introduced by E. Beyrich in 1854 for deposits classed as Miocene by some and as Eocene by others Additional studies resulted in the proposal by W Schimper, in 1874, of the term Paleocene for what had previously been considered earliest Eocene time or basal Eocene rocks

topes, both primary and secondary cosmic rays, and the neutrons resulting from spontaneous fission. The induced nuclear reactions may in turn produce projectiles for further reaction. For example, an alpha particle incident on the Al^{27} nucleus may result in its transmutation to the unstable isotope P^{30} with the emission of a neutron $Al^{27}(\alpha,n)P^{30}$; P^{30} then emits a positron decaying to the stable isotope Si^{30}. The neutron may then cause reactions, such as (n,γ), (n,α), $(n,2n)$, (n,p) or may induce fission in a heavy element. The positron is annihilated, releasing two energetic gamma rays (photons), which also may produce nuclear reactions. The net effects are the release of energy and eventually a shift in stable isotope abundances which may or may not be perceptible.

It is safe to say that virtually all reactions which can be produced artificially also occur in nature In fact, man has not yet succeeded in producing artificially the full range of naturally occurring projectiles such as the extremely high-energy nuclei in the primary cosmic ray flux However, in contrast to artificial and stellar environments, the rate at which terrestrial nuclear reactions take place is so slow as to be barely perceptible by the most sensitive techniques, and the resulting shift in stable isotope abundance is significant only for isotopes of very low abundance. For example, the stable isotope of lowest abundance, He^3, has been demonstrated to be wholly of secondary origin, resulting primarily from the spallation of atmospheric nuclei induced by cosmic radiation. In contrast, the reaction $Al^{27}(\alpha,n)P^{30}$ has shifted the terrestrial abundance of Si^{30} by about 1 part per 10^{12} during the last 3,000,000,000 years However, the total Si^{30} produced during this time, although overshadowed by the high initial abundance of Si^{30}, amounts to almost 50,000,000 tons

In general, because of their very low abundance and inert nature, the effects of nuclear reactions can most readily be observed for the rare gases. Variations of over 300% in the A^{36} to A^{38} ratio in pitchblende minerals have been ascribed by W H. Fleming and H G Thode (1953) to reactions such as $Cl^{35}(\alpha,p)A^{38}$ and $Cl^{35}(\alpha,n)K^{38} \xrightarrow{\beta^+} A^{38}$. G. W. Wetherill (1954) found variations in the Ne^{21} and Ne^{22} abundance in minerals which he ascribed to the reaction $O^{18}(\alpha,n)Ne^{21}$ and $F^{19}(\alpha,n)Ne^{22}$. T. W. Morrison and J. Pine (1955) have calculated the neutron yield from (α,n) reactions in granite (O^{18}, Na^{23}, Mg^{25}, Mg^{26}, Al^{27}, Si^{29}, Si^{30} are the most fertile targets). They were able to demonstrate that the theoretical calculation is in agreement with the directly measured neutron flux.

As previously mentioned, the neutrons released from such (α,n) reactions and from spontaneous fission, induced fission, and cosmic rays also result in nuclear reactions (for cosmic ray reactions which produce radioactive species *see* RADIOACTIVE SPECIES PRODUCED BY COSMIC RAYS). The reaction $H^1(n,\gamma)D^2$ is the result of the absorption of neutrons in all hydrogenous substances. This reaction has a practical application in well logging. He^3 resulting from the reaction $Li^6(n,\alpha)T \xrightarrow{\beta^-} He^3$ has been observed in the lithium mineral spodumene by L. T. Aldrich and A. O. Nier (1948). Transuranium elements have been separated from uranium ores (C. A. Levine and G. T. Seaborg, 1951). The total terrestrial abundance of plutonium may be estimated at approximately 100 tons. The "missing" neptunian $(4n + 1)$ series is also present in nature in minute amounts. However, the abundance of transplutonian elements seems to be too small to detect by present means.

Small amounts of He^4 have been produced by the photodisintegration process $Be^9(\gamma,n)Be^8 \rightarrow 2He^4$.

The half life for spontaneous fission is about 10^{16} years. The amount of induced fission in uranium is critically dependent upon the presence of high cross-section neutron absorbers such as the rare earths. A careful search for the existence of self-supporting chain reactions in nature made by P. K. Kuroda in 1956 indicates that, in all the older mineral assemblages studied, self-supporting chain reactions are quenched by the presence of rare earth isotopes. However, younger minerals of low rare earth content were found which might have supported a critical chain reaction if they had existed during early geologic history Natural critical chain reaction systems, formed during early earth history, were a transient phenomenon. Stable mineral assemblages of critical composition were prevented from forming or were destroyed *See* LEAD ISOTOPES, GEOCHEMISTRY OF; RADIOACTIVE MINERALS. [P.E.D.]

Bibliography: K. Rankama, *Isotope Geology*, 1954.

Territoriality

A pattern of behavior in which one or more animals occupy and defend a definite area or territory. They obtain their food from this area and exclude others of their species from it Territorial behavior is highly developed among many kinds of birds but appears also, in somewhat different forms, among mammals, reptiles, fish, and social insects. Among singing birds, the males returning in the spring migration choose territories, often the same ones occupied the year before. By singing and displaying his colors, the male advertises to female birds that a mate and territory are available. When he is joined by a female, they build a nest in the territory, and use it as a feeding area for themselves and for the young birds until these leave the nest. Meanwhile the male defends the territory by singing to warn off other birds of his species and by attacking and driving off intruders

A whole area of forest or grassland may thus be occupied by birds of the same species, each with its territory. Each singing-bird species subdivides the area in a pattern of territories that is independent of those of other species Birds of different species may move through one anothers' territories without conflict. In a forest there may be vertical

The seminiferous tubules in different species

...les. Each tubule is surrounded by a delicate membrane (Fig 2a and b) which is contractile in any species and enables the tubules to wriggle ...ly. The spaces between tubules are filled with connective tissue, blood vessels, and secretory cells, the interstitial cells or cells of Leydig, which secrete male hormone. The interstitial cells vary in different species and different periods in life. In the human fetus they are extremely abundant and presumably functional; later they regress and change into inconspicuous connective tissue cells. At puberty they reappear and enter the main productive phase of their life.

Sperm formation. The sperm cells, spermatozoa, develop in the wall of the seminiferous tubules, either periodically, as in most vertebrates, or continually, as in man. Most of the cells in the tubules are potential spermatozoa. Nursing cells (cells of Sertoli) are interspersed at regular intervals between them (Fig 2a and b).

Four cell types representing characteristic stages in the development of sperm cells have been given specific names. The youngest ones are called spermatogonia or, at their earliest stages, stem cells. They lie at the periphery of the tubule and show no resemblance to mature spermatozoa. Once a reproductive phase has begun, the spermatogonia undergo periodic divisions. A few of them stop dividing after one or two mitoses and revert to stem cells. These are the progenitors of future generations of spermatozoa. Most spermatogonia, however, go through additional divisions and transform into another cell type, the spermatocyte, in which the nucleus is slowly preparing for the maturation divisions (meiosis). The small cells resulting from these divisions are called spermatids. They are haploid, that is, their chromosome number is reduced by one-half. Usually four spermatids arise from one spermatocyte. The subsequent transformation of the spermatid into a spermatozoon is called spermiogenesis. This is an extremely complex process involving development of a "head" through condensation of the nucleus, a "head cap" at the anterior pole, and a "tail". During spermiogenesis, groups of spermatids become attached to each nursing cell and often indent its membrane deeply. Finally the spermatozoa retract toward the lumen and are released; they are still hardly motile and usually not fertile.

The duration of spermatogenesis is known with fair accuracy for several species. Usually it is about 5-7 weeks from the first spermatogonial

... than one-third in the long preparation for, and the rapid completion of the maturation divisions, and about one third in spermiogenesis.

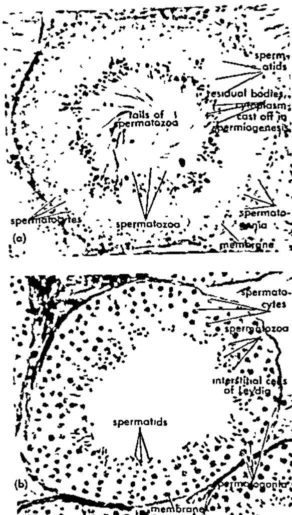


Fig 2 Seminiferous tubules in cross section. (a) Rat. (b) Human. In the rat, one particular, definite association of 4 stages of germ cells is photographed, in the human a multitude of stages is present.

In nearly all mammals, spermatogenesis occurs in rigid patterns which are essentially similar. Large groups of cells filling sections of tubules up to several centimeters in length develop synchronously. Generations follow each other at definite intervals so that in any section of a tubule, characteristic associations of cells are present (Fig 2a). In addition, along the tubule, groups of cells which are in synchronous development border on other groups in immediately preceding or following stages. This constitutes the so-called spermatogenic wave. In man and probably in some apes, the spermatogenic wave is absent or indistinct. In these species, groups of cells developing synchronously are very small and each cross section of a tubule reveals a multitude of stages of germ cells coexisting in random array (Fig. 2b).

In lower vertebrates, many patterns and arrangements of spermatogenesis are found, sometimes in gonadal sacs which open directly into the abdominal cavity (from where they are released through abdominal pores (cyclostomes), sometimes in

These divisions are now generally recognized and used throughout the world, although not strictly in accordance with their original definitions. They have been modified to some extent by additional data and by changing concepts, but even today not all geologists are fully agreed as to their precise usage and limits. In general, most modern workers do not base these divisions on percentages of living species since the percentages vary considerably from place to place in formations of the same age. Instead, subdivisions of the Tertiary are now ordinarily based on the character and stage of development of the fossils and on the stratigraphic and structural relations of the rocks containing them.

In some instances the Tertiary has been divided into two major parts: (1) an older Paleogene or Nummulitic, including Paleocene, Eocene, and Oligocene; and (2) a younger Neogene, embracing Miocene and Pliocene. Although such a procedure, at least locally, is convenient and useful, these divisions are not generally accepted or applied.

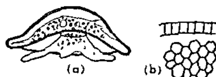
Fauna. The Tertiary fauna is dominated by mammals and birds on the land and by pelecypods, gastropods, and echinoids in the sea. During the Tertiary, archaic mammalian stocks, such as the pouched forms (marsupials) and insectivorous types, were quickly replaced everywhere except on the isolated island continents (Australia and South America). The Cretaceous-Tertiary boundary is marked by the disappearance of the large reptiles (dinosaurs, ichthyosaurs, mosasaurs, pterosaurs) and certain shell-bearing cephalopods (ammonoids) characteristic of the Mesozoic, together with the expansion of numerous mammalian stocks (marsupials, insectivores, and primitive carnivores and ungulates). Other groups of animals, particularly pelecypods, gastropods, and foraminifers, show gradual modernization from Mesozoic to Cenozoic. The Tertiary-Quaternary boundary is even more tenuous paleontologically; it is rather arbitrarily marked by the appearance of man or of works of man. The terrestrial and marine faunas of Tertiary and Quaternary times differ chiefly in distribution rather than in composition. *See* PALEOBOTANY; PALEONTOLOGY. [A.H.CH., G.E.M.]

Bibliography: *See* CENOZOIC

Testacida

An order of fresh-water Sarcodina with one-chambered tests (see illustration) and filopodia. This order is also known as the Testacea. These creeping organisms pull the body toward the point to which extended pseudopodia are attached. Major components of the test are typically siliceous which include sand grains, other foreign particles, and secreted plates. Many species thrive in acid waters of pH 5.0–6.4 which may favor extensive utilization of minerals, as in Euglyphidae.

Life cycles involve fission, in which one organism commonly builds a new test while the other receives the parental test, and encystment, or sometimes formation of a "capsule" by sealing the mouth of the test after retraction of pseudopodia.



Arcella (order Testacida). (a) Test in side view with filopodia extending ventrally, 80–140 μ in diameter. (b) Structural elements of test. (From L. H. Hyman, *The Invertebrates*, vol. 1, McGraw-Hill, 1940)

Syngamy has been reported in a few genera. *See* SARCODINA; SYNGAMY. [R.P.H.]

Testis

The organ of sperm production. In addition, the testis (testicle) is an organ of endocrine secretion in which male hormone is elaborated. In mammals, the testes are usually ovoid or round. In many species (for example, man) they are suspended in a pouch (scrotum) outside the main body cavity; in other species they are found in such a pouch only at the reproductive season; in still others the testicles are permanently located in the abdomen (for example, in whales and bats).

Histology. Within a firm and thick capsule of connective tissue, the tunica albuginea, the testis contains a varying number of thin (⅛–⅓ mm) but very long seminiferous tubules which are the sites of sperm formation. Essentially, these tubules are simple loops in the mouse and rabbit (Fig. 1, *b* and *c*) which open with both their limbs into a network of fine, slitlike canals, the rete testis. From this the sperm drains through a few, narrow ducts, ductuli efferentes, into the epididymis, the storage chamber for sperm.

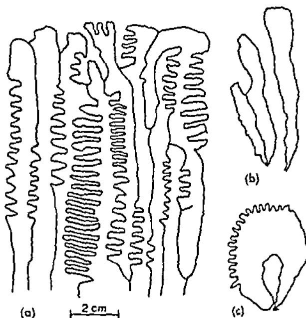


Fig. 1. Single seminiferous tubules teased out of the testis and somewhat straightened and arranged. The black lines are the hollow tubules. At their free lower ends, they were severed from their connections with the rete testis. (a) Man. (b) Mouse (after Hirota). (c) Rabbit.

their ability to increase the size and strength of bones, androgens are sometimes employed to treat osteoporosis and to promote the healing of fractures.

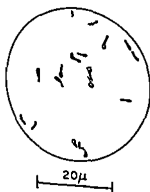
Normal pubertal development. In boys, normal

the presence of testicular hormone. Leydig cells, which are present in the prenatal testis and then disappear shortly after birth, reappear at 11–13 years. Small amounts of 17-ketosteroids are present in the urine during childhood, but they may originate from the adrenals instead of the testes. The formation of spermatozoa is usually not well established until the age of 15. With the increasing titers of testicular hormone occurring during pubescence the genital system undergoes rapid growth, and secondary sex characters, such as body hair, beard, and deeper voice, begin to appear. Psychic changes are due in part to androgens. While testicular androgen increases from ages 11-19, it is also probable that the target tissues undergo rapid maturation during this period and become increasingly sensitive to androgen. Because young infants are capable of penile erection and because intercourse with orgasm may precede by some time the first appearance of spermatozoa, it is apparent that multiple factors operate in the attainment of sexual maturity. *See* REPRODUCTIVE SYSTEM. [C.D.T.]

Tetanus

tation of deep wounds or other injured tissues. *See* TOXIN, BACTERIAL.

The causal organism is a strictly anaerobic, slender bacillus with a spherical terminal spore swelling the vegetative cell. Identification of the organism should not, however, rest on microscopic evidence but should be confirmed by neutralization



Tetanus bacilli showing terminal spores (drumsticks). Agar culture stained with fuchsin.

tests with known tetanus antitoxin of the toxin produced by the pure culture. This species produces two toxic substances, a hemolysin known as tetanolysin and the potent lethal toxin known as tetanospasmin which has a strong affinity for the cells of the central nervous system. The neurotoxin has been isolated in crystalline form. This organism does not ferment carbohydrates but depends on the fermentation of amino acids for energy. *See* TOXIN-ANTITOXIN REACTION.

The incubation period of tetanus is usually between 5 and 10 days and the disease is characterized by convulsive tonic contraction of voluntary muscles. Prevention of tetanus rests on the proper, prompt surgical care of contaminated wounds and prophylactic use of antitoxin if the individual has not been protected by active immunization with toxoid. [L.S.M.]

Tetrabranchia

A subclass of the Cephalopoda, containing the Nautiloidea and the Ammonoidea. The most numerous group of the cephalopods in the Paleozoic and Mesozoic periods, it is considered to be the most primitive group of cephalopods, originating probably as a separate line in the Precambrian. The shells of *Plectronoceras* occur in Cambrian rocks over 400,000,000 years old.

The name Tetrabranchia was proposed by R. Owen because four gills are found in *Nautilus*, the only representative whose anatomy is known. Since no records of the soft parts are preserved in fossil specimens, it is not known if this feature is shared by all of the many genera assigned to the subclass. Many workers prefer a different classification not based upon such a questionable character. *See* AMMONOIDEA; NAUTILOIDEA. [G.L.V.]

Tetractinomorpha

A subclass of the class Demospongiae in the phylum Porifera. According to recent revisions, this subclass would contain the orders Homosclerophorida, Choristida, and Clavaxinellida. These groups are diverse and constitute a heterogeneous assemblage in contrast to the other subclass, the Ceractinomorpha. *See* DEMOSPONGIAE. [C.B.C.]

Tetracycline

A broad-spectrum antibiotic. It is one of a group of antibiotics known as tetracyclines which are closely related chemically and very similar in biological action. Other well known members of the group are chlortetracycline (aureomycin) and oxytetracycline (terramycin). Tetracycline is produced biosynthetically by fermentation with a strain of *Streptomyces aureofaciens* (or certain other species) or chemically by dechlorination of chlortetracycline. Although tetracycline is not effective against the small viruses or the fungi, it is particularly useful because of broad antimicrobial action, with low toxicity, in the therapy of infections caused by gram-positive and gram-negative bacteria as well as rickettsiae and the large viruses

tubules or compartments which have only temporary connections with outleading ducts (selachians). In all lower vertebrates a multitude of spermatozoa develop in the presence of definite nursing cells. In most cases, large groups of cells, often whole compartments (in insects and amphibians) develop synchronously. See MEIOSIS; SPERMATOGENESIS. [E.C.R.R.]

Physiology. The testes are concerned with the production of the male gamete (sperm) as well as the male sex hormone, considered to be produced by the interstitial cells of Leydig which are present in the interstices between the seminiferous tubules.

Castration. Castration of prepuberal males prevents the functional development of the accessory genitalia and secondary sex characters, and delays the cessation of growth of the long bones. When the testes are removed after puberty, the libido is diminished or lost, the accessory glands involute, and certain disturbances in metabolism appear. The effects of castration vary with the species, but may be repaired by the administration of testosterone.

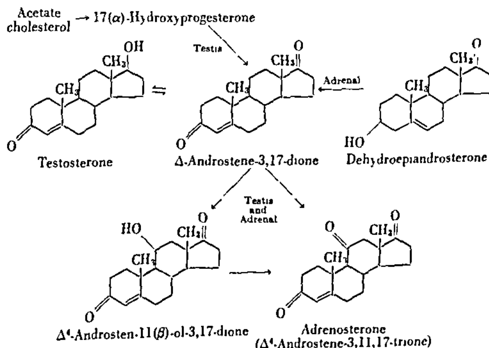
Cryptorchism. Bilateral cryptorchism, the failure of both testes to descend, invariably leads to sterility. The abnormally warm environment of the ectopic position causes destruction of the tubular epithelium. Leydig cell structure and androgen secretion may show no impairment over long periods of time. Unilateral ectopy does not lead to sterility or androgen deficiency as long as the other testis is in a scrotal position.

Hypophysectomy. Hypophysectomy of the postpuberal male stops the proliferation of sperms and reduces the secretion of testosterone to a minimum. The organism is sterile and the accessory genitalia involute to a nonfunctional state. The atrophic testes may be restored by the injection of pituitary gonadotrophins or, in certain species, by the administration of androgens. Follicle-stimulating hormone (FSH) acts to repair the tubules and promote the proliferation of sperms, but it is most effective in the presence of testosterone. Luteinizing hormone (LH) activates the Leydig cells and thus increases the output of testicular androgen. The effectiveness of LH is augmented if FSH is administered simultaneously.

Testosterone is the main androgen secreted by the testis. It has been isolated directly from testicular tissue and detected in spermatic vein blood. Androsterone is a urinary metabolite. The following scheme shows probable pathways in the biosynthesis of testicular and adrenal androgens. See ANDROGEN.

Because androsterone has a keto group at position 17 of the steroid nucleus, it may be termed a 17-ketosteroid. In alkaline solution the 17-ketosteroids give a characteristic green color with *m*-dinitrobenzene; this reaction is commonly employed in the determination of these compounds in blood and urine. During pubescence there is increased excretion of steroids of this type, and diminished excretion is noted after castration or testicular failure. The level of 17-ketosteroid excretion is an index not only of testicular function but of adrenocortical function as well. In many cases of adrenal hyperplasia, androgen production is increased and there are high titers of 17-ketosteroids in urine and blood.

The androgens are responsible for the distinguishing features of the male. One of the more obvious male attributes is the rapid development of the musculature after puberty. This is associated with a positive nitrogen balance and protein anabolism. Testosterone is sometimes used clinically for treating patients with poorly developed muscles. The hormone also increases the quantity of bone matrix and causes calcium retention. Because of



myces *aureofaciens* and the closely related *S. viridifaciens* and *S. feofaciens*; some mercially used cultures are ultraviolet mutants. *See* BACTERIAL GENETICS.

...entation methods which reduce the amount of tetracycline must be used since all tetra... ...ortetracycline strains produce some chlor... ...-producing strains produce some chlor...cycline. A combination of organic medium ...ients (usually low in chloride) and a se... strain of organism giving low chlortetracy... appear to give the best results. Other methods ...have been advocated are chloride-free syn... medium; chloride removal by chemical pre...tion or ion exchange; or the addition of ...ination inhibitors such as bromide, iodide, or ...yanate. *See* FERMENTATION.

...veral stages of inoculum development are usu... used as is common for other fermentation ...esses Starting with spores from slants or ...res dried on sand or in lyophil vials, one or more ...le flask stages may be used and then one or two ...culum tank stages. The final fermentation is ...ducted in tanks of 5000–15,000 gal Since the ...anism is strongly aerobic, the medium is me...nically agitated and sterile compressed air is ...wn through at a high rate. Suitable media may ...tain cottonseed flour, soybean meal or corn...ep liquor, a carbohydrate such as sucrose, and ...ually calcium carbonate to control pH during ...e early part of the fermentation period

Optimum antibiotic titers may be found in 72 ...ours and yields in excess of 1–2 g/liter have been ...ported. Recovery is accomplished by solvent ...traction from the broth followed by purification ...cluding, finally, crystallization Waste residues may be dried and used as supplements for animal ...eed formulations. *See* ANTIBIOTIC; INDUSTRIAL MICROBIOLOGY. [R E B.]

Bibliography: A. DiMarco and P Penella, The fermentation of the tetracyclines, *Progr In Industrial Microbiol.*, 1:47–92, 1959, H F Dowling, *Tetracycline*, 1955, Federal Trade Commission, *Economic Report on Antibiotics Manufacture*, 1958, H. S Goldberg (ed.), *Antibiotics, their Chemistry and Non Medical Uses*, 1959; T H. Jukes, *Antibiotics in Nutrition*, 1955; R E Kirk and D F Othmer (eds.), *Encyclopedia of Chemical Technology*, vol. 13, 1954.

Tetraethyllead

An organometallic compound that has found wide commercial application as an additive for motor fuels. The presence of small amounts of tetraethyllead markedly reduces the knocking tendencies of gasoline and permits the use of higher engine compression ratios and consequently increases engine efficiency.

The combustion of gasoline containing tetraethyllead produces deposits of lead and lead oxide along the cylinder walls. Hence, ethylene dibromide and ethylene dichloride are also added as part of the antiknock fluid to remove the lead as lead halide in the exhaust gases

The present commercial process for producing tetraethyllead involves the reaction of ethyl chloride and a lead-sodium alloy at moderate temperatures, followed by steam distillation of the product and recovery of the unused lead. Since tetraethyllead is poisonous and unstable, special precautions must be used in handling it. *See* ANTIKNOCK AGENTS, GASOLINE; INTERNAL COMBUSTION ENGINE; OCTANE NUMBER; ORGANOMETALLIC COMPOUND [M D.R.]

Tetrahedrite

A mineral with composition $(Cu,Fe,Zn,Ag)_{12}Sb_4S_{13}$ (essentially copper, iron, zinc, and silver, antimony, and arsenic sulfide), crystallizing in the isometric system Crystals are commonly in tetrahedrons, but the tristetrahedron, dodecahedron, and cube may be present. It is massive or granular. Its hardness is 3½–4 and the specific gravity varies from 4 6 to 5 1, depending on the composition. The luster is metallic and the color grayish-black; thus, in some mining localities, this mineral is called gray copper

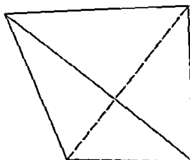
Although analyses of tetrahedrite show varying

the mineral tennantite, $(Cu,Fe,Zn,Ag)_{12}As_4S_{13}$.

Tetrahedrite is a widely distributed mineral, usually found in silver and copper veins formed at low to moderate temperatures; more rarely it is found in higher-temperature veins and in contact metamorphic deposits. It may be associated with various other copper, lead, and silver minerals, as well as pyrite and sphalerite. In some places it carries sufficient silver to make it a valuable ore, as at Freiberg, Germany, and the silver mines of Peru, Bolivia, and Mexico. It is found in various silver and copper mines in the western United States *See* ANTIMONY; ARSENIC; COPPER; SILVER. [C S.HU.]

Tetrahedron

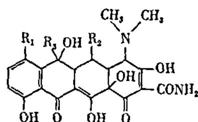
A solid bounded by four planes, or faces. It has four vertices (not coplanar), and six edges, the six line segments that join each pair of vertices. As the 3-dimensional analog of a triangle, many of its properties are extensions of those of a triangle.



Regular tetrahedron

like psittacosis-lymphogranuloma viruses. In 1958 tetracycline had domestic sales for medicinal use of over $100,000,000 compared to $20,000,000 for oxytetracycline and chlortetracycline combined. It is also used as an ingredient in animal feeds to promote enhanced growth rates, particularly in chickens and swine. *See* ACTINOMYCETALES; CHLORTETRACYCLINE; LYMPHOGRANULOMA-PSITTACOSIS GROUP; OXYTETRACYCLINE; RICKETTSIALES; STREPTOMYCETACEAE; VIRUS.

Chemical formula. Chemically, tetracycline is closely related to other members of the group. The chemical structural relationships are shown in the diagram below:



	R_1	R_2	R_3
Tetracycline	H	H	CH_3
Chlortetracycline	Cl	H	CH_3
Oxytetracycline	H	OH	CH_3
6-Demethyl tetracycline	H	H	H
7-Chloro-6-demethyl tetracycline	Cl	H	H

Tetracycline is soluble in glycol ethers, pyridine, dilute acid and alkali; it is slightly soluble in water and the lower-molecular-weight alcohols. Its stability in solution is greater than that of chlortetracycline and is comparable to that of oxytetracycline.

Antimicrobial activity. The antimicrobial activity of tetracycline is similar to that of the other tetracyclines. Usually the action is bacteriostatic, but at high concentrations, that is, 32–64 times bacteriostatic, the action may be bactericidal.

Most strains of pneumococci, β-hemolytic streptococci A, B, or C, *Streptococcus viridans, Neisseria gonorrhaeae, Hemophilus influenzae,* and *Klebsiella pneumoniae;* some strains of *Salmonella typhosa, Brucella bronchiseptica,* and *Micrococcus pyogenes* var. *aureus* (staphylococcus) are sensitive at 1 μg/ml or less. Other strains of these same species are more resistant, some to concentrations of tetracycline of over 100 μg/ml Generally more resistant are strains of enterococci, *Pseudomonas aeruginosa, Aerobacter aerogenes, Proteus* spp., *Escherichia coli, Shigella* spp , requiring usually 1–10 μg/ml or more to inhibit, but some strains are not inhibited at 100 μg/ml.

The resistance of bacterial cultures to tetracycline may be increased several hundred fold by transferring them serially to media with increasing subinhibitory concentrations of the drug. This causes similar increases in resistance to chlortetracycline and oxytetracycline (cross resistance), but not to chloramphenicol, penicillin, or streptomycin. *See* CHEMOTHERAPY; CHLORAMPHENICOL; DRUG RESISTANCE; PENICILLIN; STREPTOMYCIN.

Pharmacology. The acute toxicity of tetracycline, like that of the other tetracyclines, is relatively low. For mice the LD_{50} is 150–180 mg/kg for intravenous administration and over 3000 mg/kg for oral administration. In human therapy there are few serious toxic side reactions to tetracycline; only occasionally is an allergic reaction, nausea and vomiting, or diarrhea encountered. Some unfavorable reactions have been attributed to yeast infections following the elimination of bacterial competition; dosage forms are available which include an antifungal antibiotic to combat this condition. *See* HYPERSENSITIVITY; LETHAL DOSE 50; MYCOLOGY, MEDICAL.

Absorption Absorption of tetracycline follows oral administration and this route is usually used for therapy; intravenous and intramuscular administration is also satisfactory. Patients receiving doses of 0 25–0.5 g by mouth at 6-hour intervals have a serum or plasma concentration of 0 5–1 μg/ml in a few hours, rising to 4 μg/ml in 3–4 days where it remains constant until dosage is reduced *See* BLOOD.

Distribution. The drug readily diffuses to the cerebrospinal fluid, pleural and ascitic fluids, bile, saliva, and even across the placenta.

Excretion. About 20–25% of the dose administered orally is excreted in the urine. Small amounts appear in the milk and in prostatic fluid Over half of the administered dose is contained in the feces, some of it from bile excretion.

Therapeutic effect. A desirable therapeutic effect is obtained from doses of 1–2 g per day in pneumococcal pneumonia and in infections due to the following bacteria: hemolytic streptococci, gonococci, meningococci, and staphylococci including those staphylococci resistant to penicillin but not those resistant to other tetracyclines (because of cross resistance)

The drug is also widely used in gram negative bacterial infections such as *Shigella* dysentery, typhoid fever, *Salmonella* infections, *Hemophilus* meningitis, brucellosis, and urinary infections caused by sensitive gram-negative organisms.

Because of its broad-spectrum characteristics, tetracycline (like chlortetracycline and oxytetracycline) is used for treatment of rickettsioses and viral diseases, such as Rocky Mountain spotted fever, typhus fever, lymphogranuloma venereum, and trachoma It is without direct value in the treatment of diseases caused by the smaller viruses. *See* ANIMAL VIRUS; BACILLARY DYSENTERY; BRUCELLOSIS, GONORRHEA; MENINGITIS; PNEUMOCOCCUS; RICKETTSIOSES; SALMONELLA, STAPHYLOCOCCUS; STREPTOCOCCUS; TYPHOID FEVER.

Commercial production. Commercial production of tetracycline is by chemical synthesis, by reductive dechlorination of chlortetracycline, and also by fermentation with certain cultures of *Streptomyces* spp. The cultures used are strains of

…ters other than limbs and fins can result …rent separation For example, the eggs of …and birds possess an embryonic membrane …e amnion that permits development of the …in relatively dry situations. The same mem…lso surrounds a developing mammal and …nimals may be classified as the Amniota. …ians and other lower vertebrates lack the …; their eggs survive only in water or under …oist conditions. They are considered to com… the Anamnia. *See* AMNIOTA; AMPHIBIA; …IA, AVES, MAMMALIA; REPTILIA. [R.C.Z.]

…de, vacuum

…ode, as its name implies, is a four-electrode … The four electrodes, in the order of their …gement, are the cathode, the control grid, the … grid, and the plate. There are two types of …des, the screen-grid tube and the beam-power … Although each of these tubes has the same …ber and arrangement of electrodes, the current-…ge characteristics are quite different. This …le discusses only the screen-grid tube; *see also* …M POWER TUBE.

…he screen grid tube was developed early in the …ory of radio, when it was found that triode …lifiers had a tendency to oscillate because of … capacity feedback coupling from the plate to … grid through the interelectrode capacitance …ween these two electrodes. The screen grid is …erated at a positive dc voltage but is grounded … [illegible] With [illegible] …cillate and without the necessity for neutraliza…on previously required with triode radio-fre…uency amplifiers

These results are achieved because of the shield…g effect between the control grid and the plate …troduced by the screen grid In the ordinary …ode the grid-plate capacity is of the order of …icrofarads; in a screen grid tube it is of the order of tenths or hundredths of microfarads. The screen grid acts as an electrostatic shield between the control grid and the plate Unfortunately, the current-voltage characteristics of a screen-grid tube are not as uniform as is desired, because of the action of secondary electrons. This led to the insertion of another grid for the suppression of secondary electrons and thus to the development of the pentode.

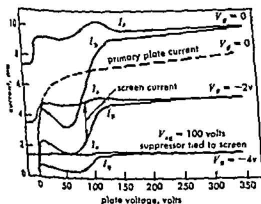


Plate-current–plate-voltage characteristics of a screen-grid tube

The illustration shows families of curves of plate current and space current as a function of plate voltage for a fixed large positive screen voltage. The dashed curve shows the form of the plate current that would be expected in the absence of secondary emission. However, when the plate is less positive than the screen grid, it will lose secondary electrons to the screen and thus reduce the plate current. When, on the other hand, the plate voltage is greater than the screen-grid voltage, the screen grid will lose secondary electrons to the plate, thus increasing the plate current. Secondary electrons are generated at all times at both the screen grid and the plate However, they can only go to the other electrode when the other is more positive.

Over this part of the plate-current characteristics the slope is negative. This corresponds to a negative resistance, and circuits containing screen-grid tubes may oscillate in this region. Because of this and because of the large distortion in the characteristics, these tubes are seldom used For further discussion, *see* VACUUM TUBE [K.R.S.]

Teuthoidea

An order of the molluscan subclass Coleoidea. Representative of the order are the squid *Loligo* and the fossil *Plesioteuthis.* The rostrum is not developed, the proostracum is represented by the elongated pen or gladius, and ten arms are present. *Loligo* is a pelagic species, frequently swimming in schools, while *Opisthoteuthis* is sedentary. Teuthoids range from the Jurassic to the Recent. *See* COLEOIDEA; DECAPODA (MOLLUSCA). [C.B.C.]

Texas Towers

Radar platforms placed on the continental shelf off the east coast of the United States, where they form a part of the Continental Air Defense System. The towers are so named because the prototype was first developed and used off the Gulf Coast of Texas for offshore oil-drilling operations The triangular-shaped platforms measure approximately 180 ft on each side and are raised 65 ft above the ocean surface. They are supported by three legs which penetrate the ocean floor.

Three such towers have been erected, known as TT-2, TT-3 and TT-4 Texas Tower No. 2, installed in 1955, is located on Georges Shoal, approximately 100 miles east of Cape Cod, Mass., in water depth of 56 ft. Texas Tower No. 3, installed in 1956, is located on Fishing Rip Shoal, approximately 25 miles southeast of Nantucket, Mass., in about 85 ft of water Texas Tower No. 4, installed in 1957, is located on Noname Shoal, about 75 miles southeast of New York City, in 185 ft. of water.

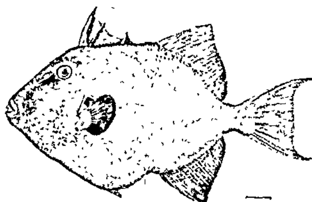
In addition to their basic military function, the towers have been equipped with facilities for geo-

Thus, as in the case of a triangle, the medians of a tetrahedron (that is, the lines joining the vertices with the centers of gravity of the opposite faces) are concurrent. On the other hand, the altitudes of a tetrahedron are not always concurrent. Although two triangles are congruent whenever the sides of one are equal, respectively, to those of the other, this is not true for tetrahedrons. There are as many as 30 tetrahedrons with the same six edges, and no two of the tetrahedrons are congruent, that is, superposable by either a rigid motion or a reflection in a plane. *See* OPTICAL ACTIVITY; POLYHEDRON. [L.M.BL.]

Tetraodontiformes

An order of specialized teleost fishes including the triggerfishes (Plectognathi), puffers, trunkfishes, and their allies. It is a group of diverse structure that is derived from typical spiny-rayed fishes (Perciformes). The posttemporal bone, if present, is simple and fused with the skull; the hyomandibular and palatine are firmly attached to the skull. The body is variably armored with bony plates or spines, encased in bone, prickly, or naked. Fin spines and some fins are variably well developed or wanting. Some species can inflate the body.

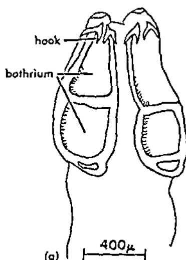
Current classifications recognize about 9 families, nearly 60 genera, and perhaps 200 species. These are largely shore fishes of tropical or subtropical seas, but a few are pelagic, enter temperate waters, or ascend tropical rivers Many are colorful inhabitants of coral reefs. Some are valued as food, but many develop alkaloids in the flesh that cause ciguatera, a frequently fatal food poisoning. *See* ACTINOPTERYGII. [R M B.]

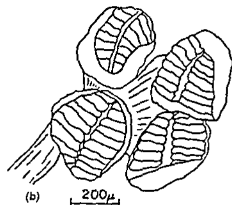


Triggerfish, Balistes carolinensis (After G B Goode, Great International Fisheries Exhibition, London, 1883, U.S. Natl. Museum Bull 27, 1884)

Tetraphyllidea

An order of tapeworms of the subclass Cestoda (phylum Platyhelminthes). All species are intestinal parasites of elasmobranch fishes and are small in size, usually less than 5 centimeters in length. An outstanding feature of the order is the variation in the structure of the holdfast organ or scolex (see illustration). All species are segmented and segments are usually shed from the body while sexually immature; these develop to sexual maturity as independent units in the host's intestine. Segment anatomy is very similar to that of Proteocephaloides A complete life cycle is not known, but larval forms have been found in a variety of invertebrates and bony fishes. *See* CESTODA; *see also* PROTEOCEPHALOIDEA. [C P.R]





(a) Acanthobothrium sp, scolex. (b) Rhinebothrium sp., scolex

Tetrapoda

A superclass of the subphylum Vertebrata that typically possesses limbs rather than fins in contrast to the other superclass of that subphylum, the Pisces

The animals comprising the Tetrapoda typically live part or all of their lives on land, whereas the members of the Pisces, the fishes, are aquatic animals The classes of the Tetrapods are Amphibia (frogs and toads, salamanders, caecilians), Reptilia (snakes and lizards, turtles, crocodiles and their kin), Aves (birds), and Mammalia (mammals) The word Tetrapoda comes from Greek words meaning "four feet," but there are tetrapods (some amphibians and reptiles) that have only two limbs or none at all. These forms have, however, evolved from four-footed ancestors.

The division of the vertebrates into the superclasses Tetrapoda and Pisces is in some respects merely a classification of convenience, and the use

on a revolving cylinder covered with very hooks or wire brushes. A moving belt, also covered with wire brushes, is on top of this cylinder. The cylinder pulls the fibers in one direction, untangles them, and arranges them in parallel in the form of a thin film (Fig 3). This film is ... molds ... sliver, ...). Card ... produces carded yarns or carded cottons suitable for inexpensive cotton fabrics. After carding, several slivers are combined. This results



Fig. 3. Inside of carding machine, where brushes clean and straighten fiber. (*Pepperell Mfg*)

in a relatively narrow lap of compactly placed staple (carded and combed) fibers; the compactness of these fibers permits this cotton stock to be attenuated or drawn out to a sliver of smaller diameter without falling apart (Fig. 5)

Combing. When the fiber is intended for fine yarns ...

... parallelism that the short fibers are combed out and separated from the longer fibers (Fig 6)

The combing process forms a comb sliver made of ...

... case. The combined process, therefore, is identified with consumers' goods of better quality. Because long staple yarns produce stronger, smoother, and more serviceable fabrics, quality cotton goods carry labels indicating that they are made from combed yarns or combed cottons.

Drawing out. The combining of several slivers for the drawing-out process eliminates irregularities that would cause too much variation if the slivers were put through singly. The drawing frame has several pairs of rollers, each advanced set of which revolves at a progressively faster speed. This action pulls the staple fibers lengthwise over each



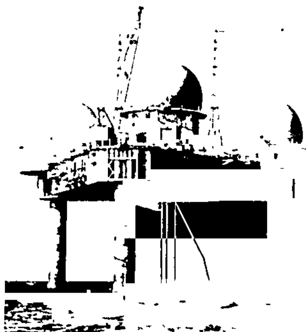
Fig. 4. Sliver leaving carding machine, where cotton has been further cleaned and disentangled. (*Pepperell Mfg*)



Fig 5. After carding, the slivers are doubled to increase the density of the future cotton yarn. (*Pepperell Mfg*)



Fig. 6. Combed slivers are combined as they leave combing machine to increase density and compactness of future yarns. Combing straightens the fibers and removes the short fibers. (*Pepperell Mfg*)



Texas Tower No. 4.

physical research. A completely equipped laboratory is in operation on TT-4 (see illustration) where oceanographic, seismic, and acoustic research is being conducted. As part of the scientific facilities of the towers, three parallel steel guide cables extend from the laboratory to the ocean bottom The cables serve as tracks for lowering and raising scientific equipment. [J.J.SC.]

Textile

A material made of natural or man-made fibers and used for the manufacture of such items as clothing, household furniture, and automobile fittings. The raw materials are fibers made of materials such as cotton, wool, nylon, acrylic resin, glass, or even metal Films and sheets made of plastic and leather are not ordinarily considered to be textiles. *See* FIBER, MAN-MADE, FIBER, NATURAL, LEATHER AND FUR PROCESSING, PLASTICS FABRICATION.

A woven fabric is constructed by interlacing or weaving sets of yarns that run lengthwise and crosswise. A close examination of any one of these y[arns] reveals the fibrous substance from which the [yarn] is made. Such yarns comprise a multitude of fi[bers] or filaments that have been separated, made p[aral]lel, overlapped, and twisted together by va[rious] processes.

There is a logical development of raw mat[erials] into consumers' goods. The products at va[rious] stages in the manufacture of fabrics from raw [ma]terial to finished goods are as follows: (1) fi[ber] is spun or twisted into yarn or directly felted [into] fabric; (2) yarn, is woven or knitted or bra[ided] into fabric; and (3) fabric, by various finis[hing] processes, is made into finished consumers' go[ods].

Spinning. The formation of yarn becomes [pos]sible when fibers have surfaces capable of cohes[ion], exemplified by the serrations of the wool fiber, [the] convolutions of the cotton fiber, and the roughn[ess] of the flax fiber. Elasticity or flexibility permits [the] fibers to be twisted around one another.

The value and character of a yarn are determi[ned] by (1) kind and quality of fiber; (2) amount [of] processing necessary to produce fineness; [and] (3) amount of twist, which increases ten[sile] strength in the finished yarn. The purpose of [the] yarn must be anticipated, because this determi[nes] the number and kind of manufacturing operatio[ns]. Important manufacturing operations in the p[ro]duction of a cotton yarn are (1) carding of lap [to] make card sliver, (2) combing of card sliver [to] make comb sliver (if the fiber is to be combed), (3) drawing-out of sliver to make roving; (4) twisting of roving to make yarn; and (5) windi[ng] of yarn on bobbins, spools, or cones.

Carding. In the lap stage the fibers are still in [a] tangled condition and contain waste material. [Be]fore this raw stock can be made into yarn, the impurities must be removed, and the fibers must [be] straightened (Fig 1) Such straightening [and] smoothing is necessary for all natural fibers; othe[r]wise, it would be impossible to produce fine yar[n] from the original tangled mass (Fig. 2). Th[e] initial process of arranging the fibers in a parall[el] fashion is known as carding The work is done on [a] carding machine on which the lap is unrolled ar[...]



Fig. 1. Cotton going through opening machine where the fibers are loosened and straightened. (*Pepperell Mfg.*)

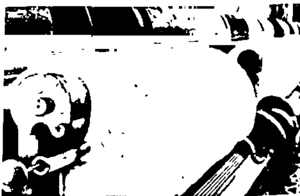


Fig 2 Cotton lap from picker room where dust[,] leaves, twigs, and other foreign matter have bee[n] removed (*Pepperell Mfg.*)

struction Compactness is one of the most significant factors affecting the durability of a fabric. It is determined by the closeness of the yarns after the fabric is woven. A closely woven fabric has a larger quantity of yarns per inch than a loosely woven one and is therefore more serviceable. A garment made from such a fabric shrinks less in washing, slips less at the seams, and is more apt to keep its shape.

A fabric of compact construction has a high thread count. Thread count, also known as cloth count, is determined by counting the number of warp yarns and filling yarns in 1 in.² of fabric. These yarns are commonly referred to as ends and picks, terms that are synonymous with warp and filling, respectively. To ascertain the thread count, a pick glass, sometimes called a thread counter, is used; this is a magnifying glass mounted on a small stand with a square opening in its base. Through this opening, filling yarns are magnified and counted.

Thread count should not be confused with yarn count. Yarn count measures the degree of fineness of yarns. Although these counts are separate devices of measurement, there is a direct relationship between them. If coarse sheeting with a low thread count is to be constructed, thick or coarse yarns will be used. These give the fabric greater resistance to hard wear.

Thread counts range from as low as 20 threads to the inch, used in tobacco cloth, to as high as [illegible] threads to the inch, found in typewriter-ribbon fabrics.

Weaving. In weaving, four operations are fundamental. They are performed in sequence and are constantly repeated. If these operations are care-

[illegible]

frame, shuttle, reed, cloth beam, and warp beam (Fig. 8). These parts perform the following operations: (1) shedding, raising warp yarns by means of the harness or heddle frame; (2) picking, inserting filling yarns by means of the shuttle; (3) battening, pushing filling yarns firmly in place by means of the reed; and (5) taking up and letting off, winding the finished fabric on the cloth beam and releasing more of the warp from the warp beam.

Shedding. On the modern loom, simple and intricate shedding operations are performed automatically by the heddle frame, a rectangular frame to which a series of wires, called heddles, are attached. As the warp yarns come from the warp beam, they must pass through openings in the heddles. Each opening may be compared to the eye of a needle. The operation of drawing each warp yarn through its appropriate heddle is known as drawing in.

In the simplest weave construction, the heddle frame raises or lowers certain groups of alternate warp yarns so that the filling yarns alternate in [illegible] under one group of [illegible]

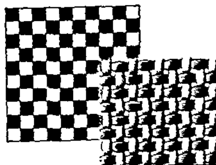


Fig 9. Construction design for plain weave resembles checkerboard, filling yarns pass under and over alternate warp yarns as shown at right. When fabric is closely constructed in plain weave, there is no distinct pattern. (From M. D Potter and B P. Corbman, Fiber to Fabric, 3d ed , McGraw-Hill, 1959)

another The heddle frame is better known as the harness, and that term is used hereafter in referring to the number of harnesses used for the different types of weaves.

Picking. As the harnesses raise the heddles, which raise the warp yarns, the filling yarn is inserted through the shed by a small carrier device called a shuttle The shuttle moves across the loom. One passage of the shuttle from one side of the loom to the other is known as a pick.

Battening. All warp yarns pass through the heddle eyelets and through openings in another frame that resembles a comb and is called a reed. With each picking operation, the reed automatically pushes, or battens, each filling yarn against the portion of the fabric that has already been formed This third essential weaving operation is therefore called battening It gives the fabric a firm, compact construction.

Taking up and letting off With each shedding, picking, and battening operation, the newly constructed fabric must be wound on the cloth beam. This process is known as taking up. At the same time, the warp yarns must be released from the warp beam; this process is referred to as letting off

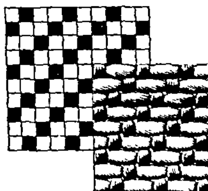


Fig 10. Three-shaft twill two warp yarns are interlaced with one filling yarn to form diagonal *(From M D. Potter and B. P Corbman, Fiber to Fabric, 3d* [illegible]



Fig 7. On spinning frame, roving passes from top through series of rollers that draw out cotton into thread. Thread is twisted as it winds onto bobbin. (Pepperell Mfg.)

other, thereby producing longer and thinner slivers. After several stages of drawing out, the condensed sliver is taken to the slubber, where rollers similar to those in the drawing frame draw out the cotton further. Here the slubbing is passed to the spindles, where it is given its first twist and is then wound on bobbins (Fig. 7)

Roving. These bobbins are placed on the roving frame, where further drawing out and twisting take place until the cotton stock is about the diameter of a pencil lead There are two stages of roving, intermediate and fine The operations are identical, but each machine yields a finer product than the stock it received. Roving has no tensile strength; it will break apart easily with any slight pull. The roving, on bobbins, is placed in the spinning frame, where it passes through several rollers running at successively higher rates of speed until it is a drawn-out yarn of a desired size.

Preparation for weaving. In the weaving operation, the lengthwise yarns that form the basic structure of the fabric are called the warp. The crosswise yarns are the filling, also referred to as the weft or the woof. The filling yarns undergo little strain in the weaving process. In preparing them for weaving, it is necessary only to spin them to the desired size and give them the amount of twist required for the type of fabric for which they will be used.

Yarns intended for the warp must pass through such operations as spooling, warping, and slashing to prepare them to withstand the strain of the weaving process. These operations do not improve the quality of the yarn. In spooling, the yarn is wound on larger spools or cones, which are placed on a rack called a creel. From this rack, the yarns are wound on a warp beam, which is similar to a huge spool. This produces an uninterrupted length of hundreds of warp yarns, all lying parallel to one another. These yarns are unwound to be put through a starch bath in a process called slashing or sizing. The slasher machine covers every yarn with a starch coating to prevent chafing or breaking during the weaving process. The yarns are passed over a large copper drum, heated by steam to set the sizing, and then wound on a final warp beam ready for the loom.

Amount of twist. The amount of twist per inch determines the appearance as well as the durability and serviceability of a fabric. Fine yarns require more twist than coarse yarns. Warp yarns are given more twist than are filling yarns. The amount of twist also depends on the type of fabric to be woven.

Yarn count. In the spinning process, there is always a fixed relation between the weight of the original quantity of fiber and the length of the yarn produced from that amount of raw material. This relation indicates the thickness of the yarn. It is determined by the extent of the drawing-out process and is designated by numbers, called the yarn count.

Thread count. The durability of a fabric depends on (1) the kind and quality of the fiber, (2) the tensile strength of the yarn, (3) the amount of twist in the yarn, (4) the use of ply yarns as compared with singles, and (5) compactness of con-

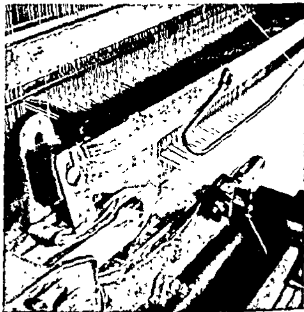


Fig 8. Action of shuttle during weaving as seen by high-speed photography Some harness frames have been raised and others lowered to form the shed or funnel of threads through which the shuttle has just passed toward the left, leaving a filling pick in its wake. The reed through which the threads are seen to pass has begun its movement forward to the solid cloth on the right, beating up the pick that the shuttle has just delivered. The reed will pound this pick firmly against the cloth It then swings back to the left, and the harness frames in the upper left-hand corner change position to form a new shed of threads, and this time the shuttle will be shot from the left side of the reed to the shuttle box on the right-hand side, leaving a new pick as it goes. (The Wool Bureau, Inc)

[illegible]poses. Finishing enhances the appearance of [th]e fabric and also may add to its serviceability [an]d durability, thus increasing its value.

Finishing may take many forms, because it must [be] adapted to the kind of fiber and yarn used in [th]e fabric and, most important of all, to its in[ten]ded purpose.

Even the factor of thread count, so important in [th]e evaluation of a fabric, can be changed by the [ki]nd and amount of finishing. Cotton can be given [th]e soft finish required for such fabrics as batiste, [n]ainsook, and lawn; the napped finish required for [fl]annelette and duvetyn; the hard stiffened finish [ty]pical of cambric and linene; or the lustrous ef[fe]ct of chintz. For discussions of important finishing [op]erations, *see* BLEACHING; DYEING; TEXTILE [CH]EMISTRY; TEXTILE PRINTING [M D P]

Bibliography: Z. Bendure and G. Pfeiffer, *America's Fabrics*, 1946; M. D. Potter and B. P. Corbman, *Fiber to Fabric*, 3d ed., 1959.

Textile chemistry

The applied science of textile materials. Textile chemistry may be considered the practice of modifying textiles chemically for wider and better use [i]n a constantly changing market.

Commercially, textile chemistry is an integral part of fabric finishing. It may involve diverse and complex phases of chemistry. It may also be concerned with engineering and chemical processing. For instance, the chemistry of a fiber is understood [m]ainly through the behavior of the fiber during chemical processing.

In recent years, especially with the advent of [illegible]

[illegible] assistants, for example, resulted in brighter and longer lasting colors. These chemicals have played a prominent role in the development of the textile industry, and they rank among the more important achievements of industrial science. The textile industry has become the greatest consumer of industrial chemicals, surpassing even the oil industry. Some 6500 chemical specialties are now manufactured for the textile industry.

Of the many varieties of fibers, only a few are [i]n largescale use: cotton, wool, flax, hemp, rayon [(vis]cose and acetate), and more recently, such [sy]nthetic fibers as nylon, orlon, and the acrylics. These have achieved preeminence because they are [ab]undant, cheap, and have desirable properties.

The textile chemist is concerned with the fabric and its performance. Performance comprises the [dur]ability, strength, dyeability, and ease of manu[f]acture of the fabric. For instance, a fabric must [m]eet certain standards of strength and resistance [to a]brasion, must resist creasing, and [illegible]

[illegible] also important to prevent creasing. [A fa]bric will be a good insulator if the fibers hold [and] trap air; wool is the best example of this. The length, strength, elasticity, and stiffness of a fiber, plus its surface characteristics, determine the ease with which the yarn can be processed, an important consideration in manufacturing costs.

Textile performance is governed by the nature of the fiber, which, in turn, is determined by the molecular structure. All fibers are made up of long-chain molecules known as polymers. The physical properties of fibers depend on the way these chains are oriented to each other and on the amount of molecular attraction among them. Polymeric fibers demonstrate crystalline behavior in x-ray photographs. Crystallinity is associated with molecular orientation and, therefore, with fiber strength. However, highly oriented fibers usually make stiff and unstretchable yarns and fabrics. Amorphous materials (materials having a limited amount of crystallinity) are generally more pliable and stretchable, but these have slow elastic recovery and less tensile strength, especially when damp. However, an amorphous material such as glass may be made into the strongest tensile fiber by drawing out the amorphous structure, while the glass is hot, until it becomes highly oriented. The manufacturer now exercises some control over orientation, crystallinity, and other factors in the polyacrylic fibers, nylon, and viscose. *See* FIBER, MAN-MADE; FIBER, NATURAL.

The textile chemist has attempted to overcome by chemical means the deficiencies of some natural fibers. For example, cotton has poor resistance to mildew, fungi, creasing, and weather. Weather resistance has been improved by copper naphthenate and other finishes. Most types of finish that chemically modify cotton have proved to be of limited commercial value because of cost, although the results are promising. Chemical processes have been devised to overcome the tendency of wool to shrink and felt (mat) in processing, but none has attained complete success.

TEXTILE AUXILIARIES

Most of the textile chemical specialties used today are substances known as auxiliaries. They increase the efficiency of the boil-off, bleaching, and dyeing operations, as well as of printing and other finishing processes. They also improve the hand (the feel), quality, and final appearance of the textile. By far the greatest number of auxiliaries is employed in the processing of natural fibers. Synthetic fibers also require auxiliaries, especially during processing. However, their action is mostly to minimize static, which is harmful in processing. Other auxiliaries give fibers the best frictional values for spinning and weaving. *See* DYEING.

The first textile auxiliary used was alkylated naphthalene sulfonate (Nekal). Since then, auxiliaries have been greatly multiplied in number; for example, there are now some 1000 brand-name detergents on the market. They are used for wetting, penetrating, dispersing, emulsifying, scouring, and other purposes. The two largest groups of auxiliaries are synthetic detergents and finishing agents.

Classification of weaves. The manner in which groups of warp yarns are raised by the harnesses to permit the insertion of the filling yarn determines the pattern of the weave and, in large measure, the kind of fabric produced. Weave patterns can create varying degrees of durability in fabrics, adding to their usefulness and also to their appearance. In a simple weave construction, consisting of the filling going under one warp and over the next, two harnesses are needed, one to lift the odd-numbered warp yarns, and a second to lift the even-numbered warp yarns. For advanced weaves, more than two harnesses are required, and as many as 40 for figured weaves.

The three basic weaves in common use for the majority of fabrics are plain, twill, and satin, with respective variations. Important constructions are also obtained from more elaborate weaves, such as pile, double cloth, gauze, swivel, lappet, dobby, and Jacquard.

Plain weave The plain weave is the simplest type of construction and is consequently inexpensive to produce (Fig. 9). On the loom, the plain weave requires only two harnesses Each filling yarn goes alternately under and over the warp yarns across the width of the fabric On its return, the yarn alternates the pattern of interlacing. If the yarns are close together, the plain weave has a high thread count; the fabric is therefore firm and will wear well.

Because the manufacture of the plain weave is relatively inexpensive, it is used extensively for cotton fabrics and for fabrics that are to be decorated with printed designs, because the surface that it produces is receptive to a direct print. The appearance of the plain weave may be varied by differences in the closeness of the weave, by different thicknesses of yarn, or by the use of contrasting colors in the warp and filling. The last method gives the effect of a design. In addition, two variations of the plain weave afford simple decorative effects, namely, the basket weave and the ribbed, or corded, weave.

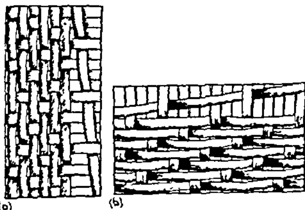


Fig. 11. Long floats typical of sateen and satin weaves. (a) Five-shaft sateen construction with floats in filling direction. (b) Eight-shaft satin construction with floats in warp direction. (*From M. D. Potter and B. P. Corbman, Fiber to Fabric, 3d ed., McGraw-Hill, 1959*)

Twill weave. A distinct design in the form of diagonals is characteristic of the second basic weave, called the twill (Fig. 10). Changes in the direction of the diagonal lines produce variations, such as the herringbone, corkscrew, entwining, and fancy twills. Increased ornamentation may be obtained by varying the diagonal, but the chief values of the twill weave are its strength, firmness, and drapability. The yarns are usually closely battened to make an especially durable fabric. Twill weaves are therefore commonly used in men's suit and coat fabrics and for work clothes, where strong texture is essential.

Satin and sateen weaves. In basic construction, the satin weave is similar to the twill weave. However, the diagonal of the satin weave is not visible because it is purposely interrupted. A continuous diagonal would interfere with the luster and smoothness desired in the satin weave.

The satin weave employs a minimum of five harnesses because a smaller number would simply form a twill weave. The use of five harnesses produces a five-shaft construction, that is, the filling yarn passes over one warp yarn and under four warp yarns. The yarn advances more than one warp yarn on each pick, thus interrupting the diagonal. When more harnesses are used, more filling yarns are interlaced by the warp yarn. The number may run as high as 11, making what is termed a 12-shaft construction because one filling yarn interlaces every twelfth warp yarn (Fig. 11).

No surface design is visible on satin fabric because the yarns that are to be thrown to the surface are greater in number and finer in count than the yarns that form the undersurface of the fabric.

Because the interlacing yarn passes over more yarns than it passes under, long yarns, called floats, are exposed on the surface of the fabric. Since these floats lie compactly on the surface with very little interruption from the yarns going at right angles to them, reflection of light on the floats gives satin fabric its characteristic luster, which is the primary object of the satin weave When the floats are in the warp, the satin weave produces a warp-faced fabric, and the luster appears in the direction of the warp. When the manner of interlacing is reversed so that the filling yarns are thrown to the fabric, the yarns will snag, roughen, and show signs of wear. But the long float gives satin fabric its desired luster and smoothness. The luster makes the fabric suitable for dressy wear, and the smoothness makes it suitable for use as lining.

Satin-weave fabrics drape well because the weave is heavier than the twill weave, which, in turn, is heavier than the plain weave. More harnesses are employed in the satin weave, thus compressing a greater amount of fine yarn into a given space of cloth Their drapability makes satin fabrics preferable for evening dresses.

Finishing. Newly constructed fabric as it comes from the loom does not represent finished consumers' goods It must pass through various finishing processes that make it suitable for many different

purposes. Finishing enhances the appearance of the fabric and also may add to its serviceability and durability, thus increasing its value.

Finishing may take many forms, because it must be adapted to the kind of fiber and yarn used in the fabric and, most important of all, to its intended purpose.

Even the factor of thread count, so important in the evaluation of a fabric, can be changed by the

flannelette and duvetyn; the hard stiffened finish typical of cambric and linene; or the lustrous effect of chintz. For discussions of important finishing operations, see BLEACHING; DYEING; TEXTILE [CHEM]ISTRY, TEXTILE PRINTING [M D P]

Bibliography: Z. Bendure and G. Pfeiffer, *America's Fabrics*, 1946; M. D. Potter and B. P. Corbman, *Fiber to Fabric*, 3d ed., 1959.

[Te]xtile chemistry

[Th]e applied science of textile materials. Textile [ch]emistry may be considered the practice of modifying textiles chemically for wider and better use [in] a constantly changing market.

Commercially, textile chemistry is an integral [p]art of fabric finishing. It may involve diverse and [c]omplex phases of chemistry. It may also be con[c]erned with engineering and chemical processing. [Fo]r instance, the chemistry of a fiber is understood [pa]rtly through the behavior of the fiber during [ch]emical processing.

In recent years, especially with the advent of

[im]portant achievements of industrial science. The textile industry has become the greatest consumer of industrial chemicals, surpassing even the oil industry. Some 6500 chemical specialties are now manufactured for the textile industry.

Of the many varieties of fibers, only a few are in large-scale use: cotton, wool, flax, hemp, rayon (viscose and acetate), and more recently, such synthetic fibers as nylon, orlon, and the acrylics.

[a]nd its performance. Performance comprises the durability, strength, dyeability, and ease of manufacture of the fabric. For instance, a fabric must meet certain standards of strength and resistance to abrasion, must resist creasing, and must wash or dry-clean without damage to color or shape. If it is to be an item of clothing, a fabric must be able to stretch and recover its shape. The property of elasticity is also important to prevent creasing. A fabric will be a good insulator if the fibers hold [...] and trap air; wool is the best example of this. The length, strength, elasticity, and stiffness of a fiber, plus its surface characteristics, determine the ease with which the yarn can be processed, an important consideration in manufacturing costs.

Textile performance is governed by the nature of the fiber, which, in turn, is determined by the molecular structure. All fibers are made up of long-chain molecules known as polymers. The physical properties of fibers depend on the way these chains are oriented to each other and on the amount of molecular attraction among them. Polymeric fibers demonstrate crystalline behavior in x-ray photo-

unstretchable yarns and fabrics. Amorphous materials (materials having a limited amount of crystallinity) are generally more pliable and stretchable, but these have slow elastic recovery and less tensile strength, especially when damp. However, an amorphous material such as glass may be made into the strongest tensile fiber by drawing out the amorphous structure, while the glass is hot, until it becomes highly oriented. The manufacturer now exercises some control over orientation, crystallinity, and other factors in the polyacrylic fibers, nylon, and viscose. See FIBER, MAN-MADE; FIBER, NATURAL.

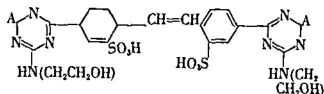
The textile chemist has attempted to overcome by chemical means the deficiencies of some natural fibers. For example, cotton has poor resistance to mildew, fungi, creasing, and weather. Weather resistance has been improved by copper naphthenate and other finishes. Most types of finish that chemically modify cotton have proved to be of limited commercial value because of cost, although the results are promising. Chemical processes have been devised to overcome the tendency of wool to shrink and felt (mat) in processing, but none has attained complete success.

TEXTILE AUXILIARIES

Most of the textile chemical specialties used today are substances known as auxiliaries. They increase the efficiency of the boil-off, bleaching, and dyeing operations, as well as of printing and other finishing processes. They also improve the hand (the feel), quality, and final appearance of the textile. By far the greatest number of auxiliaries is employed in the processing of natural fibers. Synthetic fibers also require auxiliaries, especially during processing. However, their action is mostly to minimize static, which is harmful in processing. Other auxiliaries give fibers the best frictional values for spinning and weaving. See DYEING.

The first textile auxiliary used was alkylated naphthalene sulfonate (Nekal). Since then, auxiliaries have been greatly multiplied in number; for example, there are now some 1000 brand-name detergents on the market. They are used for wetting, penetrating, dispersing, emulsifying, scouring, and other purposes. The two largest groups of auxiliaries are synthetic detergents and finishing agents.

One of the latest improvements in textile detergents and soaps is optical bleaches or whitening agents which make laundered fabrics whiter. These chemicals were developed originally from stilbene derivatives. The most advanced type is the derivative containing an *s*-triazine such as that shown in the formula, where A indicates amino residues.



The amine may contain replacement groups such as sulfo or alkoxy. Inclusion of brighteners in wash formulations gives the dried fabrics increased brightness. These agents have fluorescent characteristics which show up strongly in ultraviolet light. They also show up dyes well by increasing the color brightness.

Synthetic detergents. Detergents are substances that lower the interfacial tension of aqueous solutions. They contain a polar hydrocarbon chain of variable length plus a solubilizing portion. The hydrocarbon acts chiefly as a hydrophobic group, whereas the solubilizer is hydrophilic. *See* SOAP AND DETERGENT.

Surface-active agents. Surfactants, as surface-active agents are sometimes called, differ from one another in both the length of the polar groups and the molecular constituency of the solubilizing group, a variation that gives them a range of properties The aliphatic polar portion is usually unbranched, while the aromatic polar portion can be branched from the nucleus. Examples of solubilizing or hydrophilic groups are hydroxyl (OH), carboxyl (COOH), bisulfate (OSO_2OH), and sulfonic acid [$S(\rightarrow O)_2OH$], as well as others containing the actual group construction

Surfactants may be classified in two distinct groups, anion-active and cation-active, according to the action of the component that confers the wetting or emulsifying effect on the product For example, the Sapamines,



$$C_{17}H_{33}CONHCH_2CH_2N(CH_2CH_3)_2$$

are cationic and are effective in both acid and neutral solutions. On the other hand, the fatty alcohol sulfates of the Gardinol type and Ingepon A, $C_{17}H_{33}COOCH_2CH_2OSO_2O^-Na^+$, are anionic where the fatty chain contains the anion

This ionic characteristic is necessary because surfactants may have to operate in aqueous media of different pH values, according to the textile operation. There are now in existence a number of surface-active agents which are nonionic and can be used at a wide range of pH.

Scouring agents. Scouring is the process employed to rid the textile of foreign matter which would restrict subsequent processing. Because impurities are much more common in natural than in man-made fibers, scouring is important for these fibers.

Cotton is scoured by means of alkalies to remove noncellulosic components adhering to the fibers. Several hours of boiling are required. Detergents of alkyl aryl sulfonate types are taking the place of alkalies and may be added during soaping-off and dye-leveling operations.

Raw wool requires more attention to scouring than cotton because of the fatty nature of the substances to be removed. In the past, removal of foreign matter was accomplished by alkali soap solution, and too frequently wool damage occurred. The latest methods use synthetic detergents of the sulfated alcohol type, where there is little alkali present. Further improvements in raw-wool scouring have been made by the use of nonionics of the alkylphenol-ethylene oxide type which have gained wide acceptance, especially because alkaline solutions are avoided. Because this detergent can be used in neutral liquor, scouring and dyeing can be carried out continuously.

Silk scouring removes the silk gum (sericin) to leave the silk fiber (fibroin) unaffected. Alkali, enzymes, and new detergents have been used. As with wool, the fiber is damaged by solutions th are too alkaline.

Sizing agents. Cotton and rayon are sized protect the yarn against physical injury in t weaving operation. Cotton is immersed in star solution, and rayon is treated with gelatin a dried For some synthetic fibers, soluble polym "starches" such as polyvinyl alcohol are used. Su sizes can be rapidly removed. Sizing is sometim called slashing.

Starch sizes are removed by treatment with d sizing agents. The two main types are enzymat and chemical, such as nonionic detergents. F each type, there is an optimum pH. Gelatin siz are removed by the use of Gelatase at pH 6 0–9.0

Bleaching agents. The objective of bleachir is to produce the optimum white with the lea damage to the material. The solution pH valu vary for different fibers. In some countries, cotto is bleached in a batch process with hypochlori solutions, in the United States a continuous pro ess has been developed in which hydrogen peroxid is used Cotton is sensitive to degradation by the agents. *See* BLEACHING

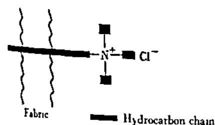
Sulfur dioxide, sulfite solutions, or peroxide ar used to bleach raw wool Whereas for cotton a pl of 10–11 is desired, wool is bleached at 7.2–7.4 Acetate rayon is alkali-sensitive; bleaching wit hypochlorite is carried out at pH 8 5–9 5. Pyro phosphate and silicate buffers may be used bot for cotton and wool; with acetate, a sulfated alco hol must be employed. Viscose and cuprammoniun rayon may be bleached with dilute hypochlorite a pH 9 4–9.8, using silicate and sulfated alcohol. So dium perborate and a sulfonated detergent have been used successfully at a pH as high as 10.2.

Finishing agents. Finishing agents can impart to fabrics a range of properties from stiffness to softness to the touch. Most of the finishing agents are used on cotton, wool, and rayon fabrics. Mer-

ızation of cotton is one of the oldest types of ‹hing process In 1851, J. Mercer discovered ıt a 15-20% caustic soda solution gave cotton a r *appearance* and properties. It became more e-absorptive, more lustrous, and stronger. Mil- ns of yards of yarn and cloth are treated by this ocess annually Cotton is immersed in solutions approximately 20% caustic while under tension. ie more fibers that are in tension the greater will the lustering effects. Therefore, the best results e obtained with the fine yarns or with heavier ırns having reversed twists in the doublings. nce the advent of optical brighteners and bluing gents, it has been possible to give cotton and other bers a brighter appearance.

Textile softeners. In search of another but elu- ıve property, fabric softness, textile chemists have chieved many different fabric effects. There are any hundreds of chemicals classed as softeners n the market. Sulfated oils and tallows are the ost common, producing soft, silky hand These inishes display ionic characteristics according to heir molecular structure. However, quaternary- ype softeners such as di-fatty acid dimethyl qua- ternary ammonium chloride are coming into popu- larity, some are employed to overcome harshness

pH range. In addition, fabrics treated with it do not turn yellow when used at high temperatures, a fault of other softeners The hydrocarbon chain, or fatty part, orients into the fabric, forming a strong



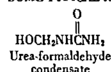
anchor The solubilizing group so anchored at the surface thus imparts soft hand to the textile. Be- cause the most common fabrics are negatively charged, cationic softeners with their positive charge have the greatest permanence.

Another way in which any degree of softness and hand can also be achieved is by the addition of synthetic fibers to natural fibers, such as viscose and cotton. Synthetic furs have been made with blends such as Dynel, Acrilan, Orlon, and Vicara to simulate natural furs. Blending has become a scientific art and is carried out in a number of processing phases such as carding, spinning, or weaving.

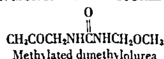
End use finishes. Besides those used to bring about more efficient processing, better hand, and better appearance, there are many types of finish which impart abrasion resistance, water repellency and waterproofness, shrink resistance, mildew re- sistance, and crease resistance.

Each of these finishes has been specially de- veloped both for the end use and the type of fiber. By far the most extensively used are those which improve ease of care in textiles made of rayon and cotton. The factors which affect fabric stability are complex. For rayon, the choice of the correct chem- icals, their processing into fiber, and the durability of the final product are some of the problems con- fronting the textile chemist. Cotton fabrics need improvement in crease resistance. The cotton fiber is highly crystalline, and its polymer structure can take only limited chemical modification. Both [illegible], can be [illegible] in fact, [illegible]inishing [illegible]

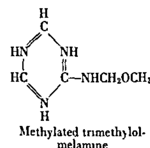
SOME TYPICAL RESIN MONOMER STRUCTURES



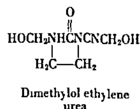
Urea-formaldehyde condensate



Methylated dimethylolurea



Methylated trimethylol- melamine



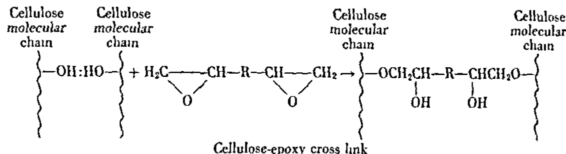
Dimethylol ethylene urea

Those resins used earliest were the urea-formal- dehyde resins, produced as shown by the reaction of urea with formaldehyde. When added to aqueous solutions, thermal polymerization insolubilizes and binds them onto the fabric. The size of molecule developed from the dispersions is both time- and *temperature-dependent. Melamine types* are more durable than the urea types, but the former retain chlorine, and cellulose fibers treated with them are liable to degradation in household use. A third resin, with the methylated urea structure, has bet- ter storage life while giving the same crease re- sistance.

A series of polymers designed for textile ease of care and durability is in development, some have reached the marketing stage and achieved wide ac- claim The first of these were the soluble epoxy resins, part of a large group of glycidyl ethers, or polyols. At present, their cost makes them com- mercially unfeasible, although they have high re- sistance to laundry degradation and damage by chlorine bleaches, as well as *freedom from* objec- tionable odors.

There is evidence that epoxy forms a cellulose- epoxy cross linking, which results in reduction of swelling and wrinkling, and greater crease re- sistance, as well as resistance to the usual cellulose

solvents. Although much secrecy exists as to the actual resin types used, there are indications that triazines and triazones are being used in some successful finishes. The latter has no keto group; the former is perhydroketotriazine. These resins are reacted with formaldehyde to make them thermosetting and fiber reactive.



Cellulose-epoxy cross link

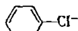
(II) shows how the silanic hydrogen cross links during oxidation. The hydrophobic methyl groups are oriented out from the fiber and the siloxane structure Si—O—Si orients toward the fiber surface forming a molecular anchor, the principle being the same as in the softening agents.

Flame resistance. Although extensive research work has been undertaken to produce flame-resistant finishes, no completely satisfactory method has become available However, one of the more successful treatments uses phosphoric acid and urea, followed by high-temperature curing, while another uses tetrakis (hydroxymethyl) phosphonium chloride.

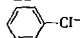
Water repellents There are numerous coating-type finishes to impart certain properties such as water resistance. All such finishes alter the surface of fibers. Some coatings are nondurable, for example, the paraffin waxes made wash-resistant by the incorporation of aluminum and zirconium salts The more durable finishes consist of highly insoluble stearic acid derivatives such as methylol stearamide, in the solubilized or dispersion form After application, they are cured at high temperatures.

DURABLE WATER REPELLENTS

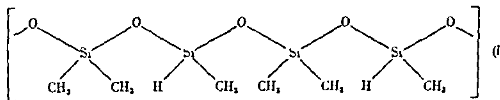








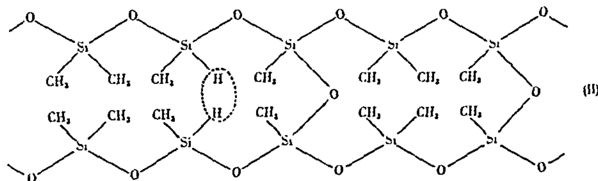
Silicones represent a newer type of water-repellent finish The principle is different from that employed in the preceding finishes because no fatty groups are involved. Instead, a plastic film enwraps the fibers. Formula (I) shows the prepolymer prepared from the hydrolysis of dimethyl dichlorosilane and methyl hydrogen dichlorosilane. Formula



NONWOVEN FABRICS

These types of textile are still chiefly in the development stage, rather than in full production. Nonwovens offer high elasticity and resiliency and can be used to back up other fabrics. They are used in shoes and surgical dressings and are finding a widening market. Cellulosics such as acetate rayon and synthetic are the fibers most often used. The fiber resins most frequently employed are polyvinyl acetate, nitrile rubber, acrylic ester, and GR-S. In some cases, thermosetting resins are desirable to cement the fibers for the particular end use.

Actually, felting may be regarded as a nonwoven material, and, therefore, pressed felt is one of the oldest forms of nonwoven fabric. This type of fabric is built up of the interlocking of fibers and requires no bonding agent The fibers become stably



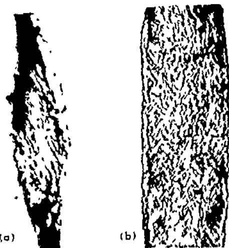
intermeshed by a combination of mechanical work, chemical action, moisture, and heat.

Another type of nonwoven material with especially short fibers 1/32–1/8 in. long is flock. The fibers are sprayed onto surfaces coated with an adhesive. Considerable quantities of flock are used in novelty items, such as wallpapers and record-player turntables. See TEXTILE. [C.C.]

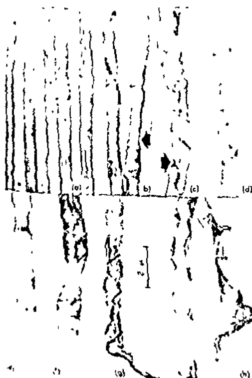
Bibliography: W. H. Cady (ed.), *Technical Manual of Am. Assoc. of Textile Chemists and Colorists*, 1958.

Textile microbiology

That branch of industrial microbiology concerned with textile materials. Most of the microorganisms on textiles—the fungi, actinomycetes, and bacteria—originate from air, soil, and water. Some of the microorganisms are harmful either to the fibers or to the consumer. They may decompose the cellulose or protein in the fiber, or affect the consumer's health. Since the minimum moisture content for microorganism development is 7%, dry storage is an effective prevention measure. Some of the microorganisms are useful, for example in the retting process, where fibers are liberated from stalks of fiber plants like flax, hemp, and jute.



Growth of cellulose bacteria (*Sporocytophaga myxococcoides* Stanier) on cotton fiber. (a) Initial stage. (b) Final stage. (A. N. J. Heyn, *Textile Research J.*, 27:591, 1957)



Stages of deterioration of cotton fiber under influence of fungi. (a) Normal fiber. (b, c) Fungal hyphae growing on the outside of the fiber. (d) Fungal hyphae growing in the lumen of the fiber. (e, f) Fibers showing extensive and irregular swelling, (e) cavitomic stage, (f) mildewed fiber with cuticle damaged and loosened (arrow). (g) and (h) Final stages of deterioration. Note fungal hyphae in spiral around fiber (g). (A. N. J. Heyn, *Textile Ind.*, vol. 120, 1956)

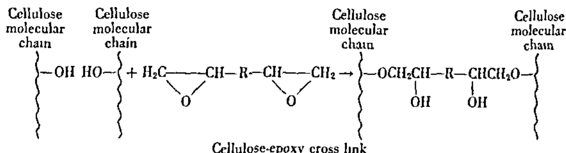
Cotton. Microbial attack of cotton can occur at any time from the opening of the boll. The fiber or fabric may be degraded resulting in a decrease in length of the fiber and strength of the cloth, uneven dyeability, darkening, or formation of colored spots. A large variety of fungi can be active in this process. Representatives of the genera *Chaetomium* and *Myrothecium* have the highest cellulose-decomposing activity; representatives of *Alternaria*, *Cladosporium*, *Fusarium*, and *Diplodia* are active in field cotton; while *Aspergillus*, *Penicillium*, and *Stachybotrys* are active in stored cotton and fabrics (mildewing). See ASCOMYCETES; MONILIALES.

The number of contaminating bacteria varies from about 1,000,000 to 100,000,000 per gram of raw cotton. Among these are the cellulose-decomposing bacteria. They all belong to the *Myxobacteriales* and have been found on field cotton where they probably play an important role in degradation. *Aerobacter cloacae* has also been found on raw cotton. This organism is of interest because it causes a respiratory disease. See MYXOBACTERALES.

Cotton is examined for microbial damage by a series of tests. In the Congo red test, microscopic examination is made of the cotton fibers after a dye, Congo red, has been applied. The undamaged cotton fibers appear pink, and the damaged cotton fibers appear red and may be cracked. In the swelling test, the degree of swelling of the cotton fibers is compared with that of normal cotton. Cotton attacked by microbes has a higher degree of swelling than normal cotton. Other testing methods determine the pH of an aqueous extract of the cotton and the fluorescence of the cotton fibers in ultraviolet light.

Wool. The number of bacteria and molds on raw wool has been reported as 1,200,000 per gram and

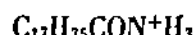
solvents. Although much secrecy exists as to the actual resin types used, there are indications that triazines and triazones are being used in some successful finishes. The latter has no keto group; the former is perhydroketotriazine. These resins are reacted with formaldehyde to make them thermosetting and fiber reactive.

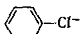


Cellulose-epoxy cross link

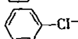
Water repellents. There are numerous coating-type finishes to impart certain properties such as water resistance. All such finishes alter the surface of fibers. Some coatings are nondurable, for example, the paraffin waxes made wash-resistant by the incorporation of aluminum and zirconium salts. The more durable finishes consist of highly insoluble stearic acid derivatives such as methylol stearamide, in the solubilized or dispersion form. After application, they are cured at high temperatures

DURABLE WATER REPELLENTS









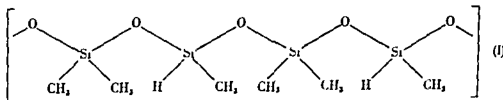
Silicones represent a newer type of water-repellent finish The principle is different from that employed in the preceding finishes because no fatty groups are involved. Instead, a plastic film enwraps the fibers. Formula (I) shows the prepolymer prepared from the hydrolysis of dimethyl dichlorosilane and methyl hydrogen dichlorosilane Formula (II) shows how the silanic hydrogen cross links during oxidation. The hydrophobic methyl groups are oriented out from the fiber and the siloxane structure Si—O—Si orients toward the fiber surface forming a molecular anchor, the principle being the same as in the softening agents.

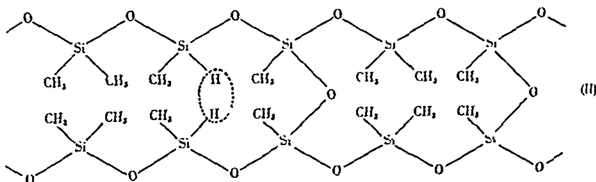
Flame resistance. Although extensive research work has been undertaken to produce flame-resistant finishes, no completely satisfactory method has become available. However, one of the more successful treatments uses phosphoric acid and urea, followed by high-temperature curing, while another uses tetrakis (hydroxymethyl) phosphonium chloride

NONWOVEN FABRICS

These types of textile are still chiefly in the development stage, rather than in full production. Nonwovens offer high elasticity and resiliency and can be used to back up other fabrics. They are used in shoes and surgical dressings and are finding a widening market. Cellulosics such as acetate rayon and synthetic are the fibers most often used. The fiber resins most frequently employed are polyvinyl acetate, nitrile rubber, acrylic ester, and GR-S. In some cases, thermosetting resins are desirable to cement the fibers for the particular end use.

Actually, felting may be regarded as a nonwoven material, and, therefore, pressed felt is one of the oldest forms of nc
ric is built up of
quires no bonding





Naphthol printing. Naphthols which are derivatives of β-oxynaphthoic acid are applied to the goods. Diazotized products called color salts are printed upon the naphtholated goods, and coupling and development takes place rapidly. The excess naphthol is then scoured out of the material.

Azoic printing. Azoic compositions, which are mixtures of the naphthols and diazotized products, temporarily inhibited from color development, are printed on the cloth. When the printed goods is passed through steam containing formic acid vapor, the coupling reaction is triggered and development takes place. The acid steaming is called acid aging.

Leuco vat esters. These are soluble reduced vat dyes stabilized in the reduced state by esterification. Acid and an oxidizing agent will cause these products to revert to their insoluble form. They are printed with an oxidizing agent and developed in the same manner as the azoic compositions above, or with a printing paste containing acid-splitting agents and developed in a neutral steaming. Acid-splitting agents are neutral salts of ammonia, for example, which can break up and yield acid on heating.

Etch printing. A material may have a body of one type of fiber and a pile of another fiber. Chemicals which attack only the pile fibers are printed on the goods, the goods is heated and the printed pile is destroyed leaving sculptured effects in the remaining pile areas which, of course, had no destructive agent printed upon them. This is also known as burn-out printing.

Flock printing. Adhesive pastes are printed on goods and while the adhesive is still wet the goods is passed between electrodes activated by alternating current. Charged particles of flock (short lengths of textile fibers) impinge upon the adhesive and remain; other areas do not permanently hold the flock. A simple blowing of the flock against the adhesive areas will also produce flock prints.

Pigment printing. This process is becoming the most widely used process in textile printing. When water is dispersed in oil, the water becomes the thickening agent. For this reason pigments dispersed in water are emulsified into solutions of resins, making successful textile pigment-printing operations possible. Simple heating of the printed goods drives off the volatile matter, and the remaining resin fixes the pigment to the goods. Newer techniques use water as the outer phase; it is claimed that sharper prints are thereby obtainable. In both types of printing paste, the pigment may be dispersed in either the water or the oil phase, but at the end of the application it is clear that the pigment is intimately involved with the resin which bonds it to the fibers.

Plissé printing. The printing of swelling agents upon the fabric enables shrinkage distortion of the whitened areas, thus imparting a plissé (crinkled) effect to the goods. Caustic soda produces this effect on cotton, and phenol similarly affects nylon. See TEXTILE CHEMISTRY. [J.E.LO.]

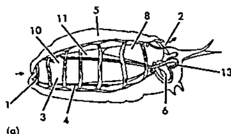
Bibliography: L. Diserens, *Chemical Technology of Dyeing and Printing*, vol. 2, 1951.

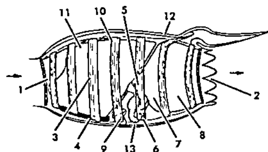
Thaliacea

A small class of pelagic Tunicata especially abundant in warmer seas. This class of animals contains three orders: the Salpida, Doliolida, and Pyrosomida. Oral and atrial apertures occur at opposite ends of the body. Members of the orders Salpida and Doliolida are transparent forms, partly or wholly ringed by muscular bands. The contractions of these bands produce currents used in propulsion, feeding, and respiration. The life cycle involves an alternation between solitary, asexual oozooids, which reproduce by budding from a complex stolon, and colonial, sexually reproducing blastozooids. The order Pyrosomida includes species which form tubular swimming colonies and which are often highly luminescent. *Salpa*, *Doliolum*, and



Colony of *Pyrosoma atlanticum*





Two examples of Thaliacea. (a) A salp, *Thalia democratica*, solitary, asexual form (after *Metcalf*); (b) a doliolid, *Doliolum*, solitary, asexual form (after *Uljanin*); (1) oral aperture, (2) atrial aperture; (3) pharynx, (4) endostyle; (5) esophagus, (6) intestine; (7) anus, (8) atrium; (9) heart, (10) dorsal ganglion, (11) muscle band, (12) tunic, (13) budding stolon

it may increase to 65,000,000 in wet, scoured wool. *Achromobacter liquefaciens*, a nonsporeformer, and various sporeforming bacteria, such as *Bacillus subtilis*, cause the greatest damage to wool. In the degradation process the fiber scales are broken down and hydrolysis of the intercellular substances of the cortex takes place. If woolen fabrics are heated to 60°C, putrefaction caused by growth of *B. subtilis* and *Agarbacterium mesentericum* may develop and produce an uneven dyeability.

Fungi may develop on stored wool under humid conditions and when degradation products of fibrous proteins are present. Microbial damage of wool can be tested, for example, by the brilliant blue fluorescence of the degreased sample under ultraviolet light, and by the absence of partial decomposition of the scales.

Retting. This is a microbial process used for liberating fiber bundles from the stalks of fiber plants. In principle, retting consists in a breaking down of pectic substances between the cell walls (middle lamellae) of the individual cells of the tissue surrounding the bundles. As a result, the bundles become separated from the surrounding tissue and can then easily be extracted mechanically.

In water retting, the stalks are immersed in cold or warm, slowly renewed water, for from 4 days to several weeks. The water source may be from rivers or constructed containers. The active organism is *Clostridium felsineum* and related types, which break down the pectin to a mixture of organic acids (chiefly acetic and butyric), alcohols (butanol, ethanol, and methanol), carbon dioxide (CO_2) and hydrogen (H_2) Its optimum activity at 33°C is the basis for the warm water retting process Related species may be active in cold-water retting In dew retting the stems are spread out in moist meadows; here the pectin decomposition is accomplished by molds and aerobic bacteria with the formation of CO_2 and H_2O. A retting process has been found to cause the liberation of fibers from the fruit husk of the coconut during soaking. *See* BACILLACEAE; INDUSTRIAL MICROBIOLOGY.

[A.N J.H.]

Bibliography: A. N. J. Heyn, *Fiber Microscopy*, 1954; R. G. H Siu, *Microbial Decomposition of Cellulose*, 1951.

Textile printing

The specialized dyeing of restricted areas on fabrics. For discussions of related topics *see* DYE; DYEING.

The application of color to textiles in printing involves the preparation of a printing paste which must contain sufficient color, the assistants to develop and fix the color, and enough thickening agent to confine the paste applications to the desired areas on the goods. The printing paste is applied to the cloth by intaglio-engraved metal rollers (machine printing) or by squeegee through negative design-coated silk screens (screen printing). The printed cloth is then subjected to heat, steam, acid fumes, or whatever will develop the locally colored areas. Final washing and drying as needed are part of the process. Because of the basic similarity to dyeing, only special types of printing will be discussed in detail.

Humectants such as glycerine, glycols, and urea are incorporated in the printing pastes to promote the solubility of some dyes but mainly to attract and hold water which is necessary to the development of the dye and to its fixation in the fiber.

In water-paste printing, dextrins, starches, gums, cellulose ethers, alginates, and emulsions are used to thicken the dye solution and to form the printing paste.

Resist printing. Reserving agents are chemicals used in the printing pastes whose function is to prevent development or fixation of the color-forming body in another printing paste or in the goods itself. For example, a red design could be printed on goods from a paste containing alkali or other chemicals. An aniline black printed as a large blotch completely over this design would not develop in the alkaline area and would produce a black area containing a red inset.

Discharge printing. This process involves printing of a color-destroying agent (usually hydrosulfite) onto dyed goods, resulting in a white design on a colored background. If the discharge paste contained a dye not destroyed by the other ingredients, then a colored design would appear on the dyed material This latter device is called color discharge printing.

Vat colors. In this process, the colors are printed from a paste containing reducing agent, alkali, and dye For best results the color must not become reduced until the goods is steamed. For this reason, reducing agents such as sodium formaldehyde sulfoxylate or thiourea dioxide are used because they do not exert their full reducing potential until high temperatures are reached. Such vat prints are fixed by a pass of about 7 min through a steam box which is called a vat ager Washing in an acidified oxidizing solution develops the color. Vat colors may also be printed in emulsions instead of in thickened pastes.

Flash aging. Because conventional vat color printing pastes tend to reduce prior to the pass through the steamer, a technique not requiring reducing agents in the paste has been developed A thickener which is practically insoluble in caustic soda is mixed with vat color. This paste is printed on the goods and dried. The goods may be developed at once or safely stored until a convenient time because no reactive substance is contained in the paste. Ultimately the printed goods is run through a padder containing caustic and hydrosulfite, and then is passed directly into hot steam for rapid development. This process requires vat dyes of specially controlled particle size to prevent incomplete development or migration of dye out of the printed area.

Naphthol printing. Naphthols which are derivatives of β-oxynaphthoic acid are applied to the goods. Diazotized products called color salts are printed upon the naphtholated goods, and coupling and development takes place rapidly. The excess naphthol is then scoured out of the material.

Azoic printing. Azoic compositions, which are mixtures of the naphthols and diazotized products, temporarily inhibited from color development, are printed on the cloth. When the printed goods is passed through steam containing formic acid vapor, the coupling reaction is triggered and development takes place. The acid steaming is called acid aging.

Leuco vat esters. These are soluble reduced vat dyes stabilized in the reduced state by esterification. Acid and an oxidizing agent will cause these products to revert to their insoluble form. They are printed with an oxidizing agent and developed in the same manner as the azoic compositions above, or with a printing paste containing acid-splitting agents and developed in a neutral steaming. Acid-splitting agents are neutral salts of ammonia, for example, which can break up and yield acid on heating.

Etch printing. A material may have a body of one [illegible] ing pile areas which, of course, had no destructive agent printed upon them. This is also known as burn-out printing.

Flock printing. Adhesive pastes are printed on goods, and while the adhesive is still wet the goods is passed between electrodes activated by alternating current. Charged particles of flock (short lengths of textile fibers) impinge upon the adhe- [illegible] prints.

Pigment printing. This process is becoming the most widely used process in textile printing. When water is dispersed in oil, the water becomes the thickening agent. For this reason pigments dispersed in water are emulsified into solutions of resins, making successful textile pigment-printing operations possible. Simple heating of the printed goods drives off the volatile matter, and the remaining resin fixes the pigment to the goods. Newer techniques use water as the outer phase; it is claimed that sharper prints are thereby obtainable. In both types of printing paste, the pigment may be dispersed in either the water or the oil phase, but at the end of the application it is clear that the pigment is intimately involved with the resin which bonds it to the fibers.

Plissé printing. The printing of swelling agents [illegible]

fect on cotton, and phenol similarly affects nylon. See TEXTILE CHEMISTRY. [J.E.LO.]

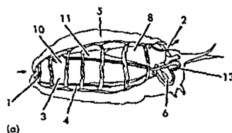
Bibliography: L. Diserens, *Chemical Technology of Dyeing and Printing*, vol. 2, 1951.

Thaliacea

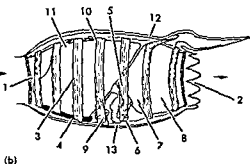
A small class of pelagic Tunicata especially abundant in warmer seas. This class of animals contains three orders: the Salpida, Doliolida, and Pyrosomida. Oral and atrial apertures occur at opposite ends of the body. Members of the orders Salpida and Doliolida are transparent forms, partly or wholly ringed by muscular bands. The contractions of these bands produce currents used in propulsion, feeding, and respiration. The life cycle involves an alternation between solitary, asexual oozooids, which reproduce by budding from a complex stolon, and colonial, sexually reproducing blastozooids. The order Pyrosomida includes species which form tubular swimming colonies and which are often highly luminescent. *Salpa*, *Doliolum*, and



Colony of *Pyrosoma atlanticum*.



(a)



(b)

Two examples of Thaliacea. (a) A salp, *Thalia democratica*, solitary, asexual form (after Metcalf); (b) a doliolid, *Doliolum*; solitary, asexual form (after Uljanin). (1) oral aperture; (2) atrial aperture; (3) pharynx, (4) endostyle, (5) esophagus; (6) intestine; (7) anus, (8) atrium, (9) heart, (10) dorsal ganglion, (11) muscle band, (12) tunic, (13) budding stolon.

it may increase to 65,000,000 in wet, scoured wool. *Achromobacter liquefaciens*, a nonsporeformer, and various sporeforming bacteria, such as *Bacillus subtilis*, cause the greatest damage to wool. In the degradation process the fiber scales are broken down and hydrolysis of the intercellular substances of the cortex takes place. If woolen fabrics are heated to 60°C, putrefaction caused by growth of *B. subtilis* and *Agarbacterium mesentericum* may develop and produce an uneven dyeability.

Fungi may develop on stored wool under humid conditions and when degradation products of fibrous proteins are present. Microbial damage of wool can be tested, for example, by the brilliant blue fluorescence of the degreased sample under ultraviolet light, and by the absence of partial decomposition of the scales.

Retting. This is a microbial process used for liberating fiber bundles from the stalks of fiber plants. In principle, retting consists in a breaking down of pectic substances between the cell walls (middle lamellae) of the individual cells of the tissue surrounding the bundles. As a result, the bundles become separated from the surrounding tissue and can then easily be extracted mechanically.

In water retting, the stalks are immersed in cold or warm, slowly renewed water, for from 4 days to several weeks The water source may be from rivers or constructed containers. The active organism is *Clostridium felsineum* and related types, which break down the pectin to a mixture of organic acids (chiefly acetic and butyric), alcohols (butanol, ethanol, and methanol), carbon dioxide (CO_2) and hydrogen (H_2). Its optimum activity at 33°C is the basis for the warm water retting process. Related species may be active in cold-water retting In dew retting the stems are spread out in moist meadows, here the pectin decomposition is accomplished by molds and aerobic bacteria with the formation of CO_2 and H_2O. A retting process has been found to cause the liberation of fibers from the fruit husk of the coconut during soaking. *See* BACILLACEAE; INDUSTRIAL MICROBIOLOGY.

{A.N.J.H.}

Bibliography: A. N. J. Heyn, *Fiber Microscopy*, 1954. R. G. H. Siu, *Microbial Decomposition of Cellulose*, 1951.

Textile printing

The specialized dyeing of restricted areas on fabrics. For discussions of related topics *see* DYE; DYEING.

The application of color to textiles in printing involves the preparation of a printing paste which must contain sufficient color, the assistants to develop and fix the color, and enough thickening agent to confine the paste applications to the desired areas on the goods. The printing paste is applied to the cloth by intaglio-engraved metal rollers (machine printing) or by squeegee through negative design-coated silk screens (screen printing). The printed cloth is then subjected to heat, steam, acid fumes, or whatever will develop the locally colored areas. Final washing and drying as needed are part of the process. Because of the basic similarity to dyeing, only special types of printing will be discussed in detail.

Humectants such as glycerine, glycols, and urea are incorporated in the printing pastes to promote the solubility of some dyes but mainly to attract and hold water which is necessary to the development of the dye and to its fixation in the fiber.

In water-paste printing, dextrins, starches, gums, cellulose ethers, alginates, and emulsions are used to thicken the dye solution and to form the printing paste.

Resist printing. Reserving agents are chemicals used in the printing pastes whose function is to prevent development or fixation of the color-forming body in another printing paste or in the goods itself. For example, a red design could be printed on goods from a paste containing alkali or other chemicals. An aniline black printed as a large blotch completely over this design would not develop in the alkaline area and would produce a black area containing a red inset.

Discharge printing. This process involves printing of a color-destroying agent (usually hydrosulfite) onto dyed goods, resulting in a white design on a colored background. If the discharge paste contained a dye not destroyed by the other ingredients, then a colored design would appear on the dyed material. This latter device is called color discharge printing

Vat colors. In this process, the colors are printed from a paste containing reducing agent, alkali, and dye. For best results the color must not become reduced until the goods is steamed. For this reason, reducing agents such as sodium formaldehyde sulfoxylate or thiourea dioxide are used because they do not exert their full reducing potential until high temperatures are reached. Such vat prints are fixed by a pass of about 7 min through a steam box which is called a vat ager. Washing in an acidified oxidizing solution develops the color. Vat colors may also be printed in emulsions instead of in thickened pastes.

Flash aging. Because conventional vat color printing pastes tend to reduce prior to the pass through the steamer, a technique not requiring reducing agents in the paste has been developed. A thickener which is practically insoluble in caustic soda is mixed with vat color This paste is printed on the goods and dried. The goods may be developed at once or safely stored until a convenient time because no reactive substance is contained in the paste. Ultimately the printed goods is run through a padder containing caustic and hydrosulfite, and then is passed directly into hot steam for rapid development. This process requires vat dyes of specially controlled particle size to prevent incomplete development or migration of dye out of the printed area.

...m pressure It decomposes at 130°C. In contrast, ...e dialkyls are very stable thermally and are re...stant to the action of oxygen and water as well. The monophenyl dichlorothallium (mp 234°C) is prepared by boiling thallium(I) chloride with ...phenylboric acid in water. It tends to decompose to the diphenyl compounds and thallium. The alkyl members of this series, like those of the other two classes, are less stable than the aryl compounds.

Analysis. Thallium may be determined spectroscopically or in solution by oxidimetry. The Tl_2O_3 may also be precipitated and carefully dried to avoid decomposition. *See* GALLIUM; INDIUM. [E.M L.]

Bibliography: J. Kleinberg (ed.), *Treatise on Inorganic Chemistry*, vol. 2, 1956; N. V. Sidgwick, *The Chemical Elements and Their Compounds*, vol. 1, 1950

...ophyta

...f the two subkingdoms of plants. It consists ...ast array of lower, less developed species as ...sted to the more advanced forms in the ...r subkingdom Embryophyta. The Thallophyta ...de numerous remotely related plants, such as

...s to large multicellular plants, some more ...200 ft long.

...espite great diversity, the Thallophyta have ...e characteristics in common: (1) none poss...es the vascular tissues xylem (water-conduct...) and phloem (food-conducting), (2) for this ...son, none has true roots, stems, or leaves, al...ugh some species have rudimentary structures ...ch resemble these organs, (3) the sex "organs" ...d the sporangia (spore-producing structures) ...usually unicellular; the gametes (sex cells) ...d spores are not enclosed in a jacket of sterile ...ll cells; (4) the zygotes (fused gametes) do not ...oduce multicellular embryos while still in the fe...ale sex "organs," as in higher plants.

Thallophyta is divided into seven phyla of algae: ...anophyta, Euglenophyta, Chlorophyta, Chryso...hyta, Pyrrophyta, Phaeophyta, and Rhodophyta, ...nd into three phyla of fungi· Schizomycophyta, Myxomycophyta, and Eumycophyta See separate articles for descriptions of each of these phyla. *See also* ALGAE; FUNGI, PLANT KINGDOM [P D S]

Bibliography: H C. Bold, *Morphology of Plants*, 1957; H. J. Fuller and O Tippo, *College Botany*, rev. ed., 1954; G M Smith, *Cryptogamic Botany*, vol. 1, 2d ed., 1955.

Thecanephria

An order of the phylum Brachiata containing a group of elongate, tube-dwelling, tentaculate, deep-sea animals of bizarre structure. The order comprises the families Polybrachiidae, Lamellisabellidae, and Spirobrachiidae. This order was erected by A. V Ivanov in 1955, to distinguish these families from the single family in the Athecanephria, the other order of the class. The anterior coelom is horseshoe-shaped. The excretory sections of the coelomoducts in the first segment are approximated medially and greatly convoluted. They are situated in the ventral, saclike invagination of the wall of the dorsal vessel. There is no pericardium. These features provide the basis for the separation of the two orders. *See* BRACHIATA. [T.H.B.]

Thecodontia

The most primitive order of archosaurian reptiles, confined to the Triassic. They differ from Lepidosauria in the absence of a supratemporal bone, parietal foramen, and palatal teeth, and in the presence of an antorbital fenestra; dorsal intercentra are absent. Primitive features in comparison to other archosaurs include retention of clavicles and interclavicle, and platelike development of pubis and ischium, which, however, show the beginning of elongation associated with bipedalism. The feet were 5-toed, and rear limbs were always longer than front limbs. The order is ancestral to crocodiles, pterosaurs, birds, and both saurischian and ornithischian dinosaurs *See* ARCHOSAURIA; LEPIDOSAURIA.

Chasmatosaurus, the oldest archosaur, with a pointed skull and down-curved snout, and *Erythrosuchus*, a heavily built quadruped 4 m long, are both from the Early Triassic of South Africa. Typical armored pseudosuchians of the Late Triassic range from the lizardlike *Aetosaurus* to *Desmatosuchus* (Fig. 1) which was more than 2½ m long

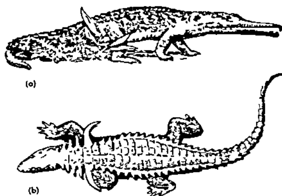


Fig. 1 Restorations of the quadrupedal thecodonts (a) *Phytosaurus* and (b) *Desmatosuchus* (From E. H Colbert, *Evolution of the Vertebrates*, Wiley, 1955)

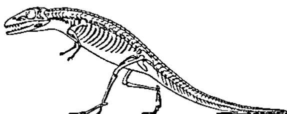


Fig 2 Skeleton of *Saltoposuchus*, a Late Triassic pseudosuchian from Germany. (After von Huene)

Pyrosoma are familiar genera. *See* TUNICATA; *see also* BIOLUMINESCENCE. [D.P.A.]

Bibliography: H. Thompson, *Pelagic Tunicates of Australia*, Commonwealth Council for Scientific and Industrial Research, Australia, 1948.

Thallium

Chemical element number 81, thallium, Tl, has atomic weight 204.39. Thallium is a member of group III of the periodic table and of the sixth period. Its valence-electron configuration is $6s^2 6p^1$ which results in a maximum oxidation state of III and a lower oxidation state of I. The Goldschmidt radius for the Tl^{3+} ion is 1.05 A, and for the Tl^{1+} ion 1.49 A.

Thallium occurs as a minor constituent in iron, copper, sulfide, and selenide ores. Minerals of thallium are considered rare. It occurs in the earth's crust to the extent of 0.6×10^{-4}%.

There are no large-scale uses for thallium metal, and indeed there is little demand for thallium compounds. Thallium compounds are toxic to humans, and other forms of life. When mixed with food, they have been used as rodenticides.

Properties of the element. Thallium is recovered from the flue dust of roasting operations. It is extracted from these residues, and recovered as the metal by electrolytic reduction from sulfate solutions. It is a white, soft metal with a melting point of 302.5°C, a boiling point of 1460°C, and a density of 11.83 g/cm³. The metal is capable of being oxidized by hydrogen ion as shown by the standard electrode potential of +0.3363 volts for the reaction



$$Tl \rightarrow Tl^{1+} + e^-$$

To dissolve the metal, however, an oxidizing acid such as nitric acid is used, because thallium chloride and sulfate are not very soluble, and their formation interferes with the oxidation. Thallium metal reacts with the halogens and oxygen at elevated temperatures to form the oxidation state I compounds.

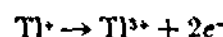
Principal compounds. Thallium(I) fluoride is soluble in water, but thallium chloride, bromide, and iodide are only slightly soluble, with the solubility progressively decreasing in that order.

Thallium(I) oxide is a black powder which reacts with water to yield the hydroxide. The hydroxide, a yellow crystalline substance, is quite soluble in water and is a strong base.

Thallium(I) halides, TlX

	Melting point, °C	Boiling point, °C
TlF	327	655
TlCl	430	806
TlBr	456	815
TlI	440	824

Thallium(I) ions can be oxidized to thallium(III) in solution by a good oxidizing agent. The standard potential for the reaction



$$Tl^{+} \rightarrow Tl^{3+} + 2e^-$$

is −1.25 volts, which is just about the same as the standard potential for the oxygen-water couple. In hot solutions, thallium(III) is unstable with respect to reduction to the I state by water. Like thallium(I), thallium(III) ions are colorless in solution.

When base is added to a solution of thallium(III), the brown Tl_2O_3 is precipitated; the hydroxide has been shown to be nonexistent. The carbonates, sulfates, phosphates, and oxalates of thallium(III) are also water-insoluble. The thallium(III) oxide starts to decompose to thallium(I) oxide at 100°C.

The halides of thallium(III) are unstable thermally with respect to the halogen and the thallium(I) halide. The trifluoride is the most stable, and can be prepared by the reaction of fluorine with thallium(III) oxide. TlF_3 melts in an atmosphere of fluorine at about 550°C, but decomposes when heated in air, and hydrolyzes in water. Oxidation of thallium(I) chloride by chlorine gives thallium(III) chloride, which decomposes upon melting at 25°C. The bromide shows an appreciable partial pressure of bromine even at room temperature. The iodide apparently is not $(Tl^{3+}3I^-)$ but rather (Tl^+, I_3^-), at least in the solid state.

Compounds containing mixed oxidation states of thallium can be obtained by careful decomposition of the thallium(III) halides. Examples of these are $Tl_3(TlCl_6)$ and $Tl(TlBr_4)$. With the alkali metal halides, hydrated salts are obtained in which the oxidation state three is stabilized as the anion, for example, $NaTlI_4$, $KTlBr_4$, K_3TlCl_6, and $Cs_3Tl_2Cl_9$. In the latter case, thallium occurs in a binuclear anion with each thallium surrounded by six chlorines octahedrally, and with three chlorines making a common face.

Thallium forms organometallic compounds of the following general classes, R_3Tl, R_2TlX, and $RTlX_2$, where R may be an alkyl or aryl group, and X a halogen. The compound Et_2TlCl is prepared by the reaction of a Grignard reagent, ethyl magnesium bromide, with thallium(III) chloride. The triethyl product Et_3Tl is prepared in turn by the reaction of lithium ethyl with the diethylthallium chloride in an appropriate organic solvent. Triethylthallium is a yellow liquid with a melting point of −63°C and a boiling point of 55°C at 1.5

Theoretical physics

The description of natural phenomena in mathematical form. *It is impossible to separate theoret-* [illegible] ... by the application of [illegible] ... See PHYSICS.

Purposes. There are two main purposes of theoretical physics, the discovery of the fundamental laws of nature and the derivation of conclusions from these fundamental laws.

Discovery of fundamental laws. Physicists aim to reduce the number of laws to a minimum to have as far as possible a unified theory. When the laws are known, it is possible from any given initial conditions of a physical system to derive the subsequent events in the system. Sometimes, especially in quantum theory, *only the probability of various* events can be predicted. *See* QUANTUM MECHANICS; QUANTUM THEORY, NONRELATIVISTIC.

Conclusions from fundamental laws. These may be of several different types.

1. Conclusions may be derived in order to test a given theory, particularly a new theory. An example is the derivation of the spectrum of the hydrogen atom from quantum mechanics; the verification of the predictions by accurate measurements is a good test of quantum mechanics (*see* ATOMIC STRUCTURE AND SPECTRA). On rather rare occasions experiment has been found to contradict the predictions of an existing theory, and this has then led to the discovery of important new physical laws. An example is the Michelson-Morley experiment on the constancy of the velocity of light, which led to special relativity theory. For a discussion of this experiment, *see* LIGHT.

2. Theory may be required *for experiments designed* to determine physical constants. Most fundamental physical constants *cannot be accurately* measured directly. Elaborate theories may be required to deduce the *constant* from indirect experiments. An example is the Millikan oil-drop determination of the *electron* charge, which requires the knowledge of the motion of small droplets in air as deduced from hydrodynamic theory. *See* ATOMIC CONSTANTS.

3. Predictions *of* physical phenomena may be made in order to gain understanding of the structure of the *physical* world. In this category fall theories of the structure of the atom leading to an understanding of the periodic system of elements, or of the structure of the nucleus in which various models are tested (shell model, collective model, etc.). In the same category fall applications of theoretical physics to other sciences, for example, to chemistry (theory of the chemical bond and of the rate of chemical reactions), astronomy (theory of [illegible] ... applications may be drawn from fundamental laws. All of engineering may be considered an application of physics, and much of it is an application of mathematical physics, such as elasticity theory, aerodynamics, electricity, and magnetism. The generation and propagation of radio waves of all frequencies is one of the relatively recent examples of application of theoretical physics to direct practice. *See* ELECTRICITY; MAGNETISM; RADIO-WAVE PROPAGATION.

Content. Apart from the classification of the fields of theoretical physics *according to purpose,* a classification can also be made according to content. Here one may perhaps distinguish three classification principles: type of forces, scale of physical phenomena, and type of phenomena.

Types of force. At present four different types of force are known to physics. The best understood is electricity and magnetism. Here the fundamental laws, Maxwell's equations, are completely known. Corrections due to quantum theory exist but can be calculated. For practical purposes electromagnetic *fields can be calculated with confidence and precision,* from dc fields to the shortest γ-ray wavelength. *See* ELECTROMAGNETIC RADIATION; FORCE; MAXWELL'S EQUATIONS.

The second type of force, known to man for the longest time, is the gravitational force. For practical purposes, Newton's inverse-square law is usually sufficient. The most complete theory of gravitation, however, is Einstein's general theory of relativity, which has great beauty but only limited experimental confirmation; therefore rival theories are sometimes proposed. *See* GRAVITATION; RELATIVITY.

The strong force which holds atomic nuclei together is the third type. In contrast to the first two types, *only some of the general features of the* nuclear force are known at present. No exact *quantitative predictions from first principles,* similar to the very successful ones in electrodynamics, can be made in nuclear theory. *It* is known that nuclear forces are related to various unstable particles (mesons), but this relation is only partially understood. The nuclear force is the strongest force known to physics but extends only over very small distances. *See* NUCLEAR STRUCTURE.

Distinct from the nuclear force are the weak forces responsible for beta radioactivity and similar phenomena. They are probably closer to being understood than the strong nuclear force.

Scale of physical phenomena. The motion of bodies on the scale of everyday life can be described by the classical mechanics of Isaac Newton. Phenomena in very small dimensions, especially inside atoms or atomic nuclei, must be described by quantum mechanics. The latter theory contains Newton's mechanics as a special case. *See* MECHANICS, CLASSICAL.

The description of physical phenomena is also different according to the *velocities of the* bodies involved. When the velocity is a substantial fraction of that of light, the *special theory of* relativity must be used to describe the motion. (The special

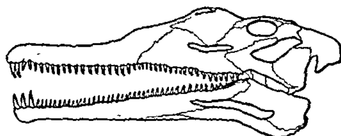


Fig 3. Skull of *Phytosaurus*. N indicates position of external nostril. (After E. H. Colbert)

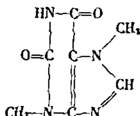
and had long horns curving over its shoulder. *Saltoposuchus* (Fig. 2) and *Ornithosuchus* had small bony plates in rows beneath the scales along the back.

The phytosaurs (Parasuchia) were Late Triassic thecodonts specialized for aquatic life in a fashion similar to crocodiles and gavials, from which they differ most obviously in the posterior position of the external nostrils, absence of a secondary palate, and pelvic and shoulder girdles. *Paleorhinus* and *Mystriosuchus* had extremely slender rostra; *Phytosaurus* (Fig. 3) was massively built. [J.T.G.]

Bibliography: A. S Romer, *Vertebrate Paleontology*, 2d ed., 1945

Theobromine

An alkaloid prepared from the dried ripe seed of *Theobroma cacao* or made synthetically. It is often extracted from waste products of the cocoa and chocolate industry. Theobromine is a close chemical relative of caffeine.

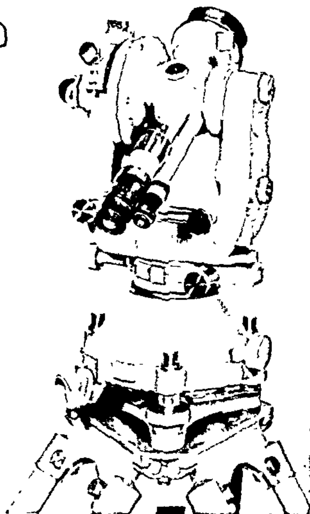


Theobromine is known to be responsible, in part, for the stimulant action of cocoa and other beverages. In addition, it is used therapeutically as a diuretic in the treatment of various types of edemas. Because of its relatively low solubility, it is used mostly in the form of mixtures which are much more soluble. *See* ALKALOID, CACAO

[S.M.K.]

Theodolite

A surveying instrument for measuring horizontal and vertical angles, similar in principle to the transit. In America the term implies maximum precision; the graduated circles for angular readings are large, and the telescope has high magnification. A sensitive level bubble on its trunnion axis prevents the telescope-reversal feature of the transit. A repeating theodolite has the transit's upper and lower motion, permitting angle repetition. A direction theodolite's horizontal circle is fixed on the leveling head, requiring the computation of angles by subtraction of the smaller direction reading from the larger direction reading. In Europe, theodolite commonly refers to any type of precision surveying instrument for horizontal and vertical angle measurement *See* SURVEYING; TRANSIT (ENGINEERING) [R.H.DO.]



A first-order theodolite. (*Wild Heerbrugg Instrument Co*)

Theorem

A proposition arrived at by the methods of logical deduction from a set of basic postulates or axioms accepted as primitive and therefore not subject to proof. Since a theorem is part of a purely formal system it is not meaningful to speak about the "truth" of a theorem but only about its "correctness." That is, the only point at issue with regard to a theorem is whether there is any flaw in the logical steps by which it has been deduced from the postulates. The classic example of a system of theorems is afforded by Euclid's system of geometry, which is now recognized to have only a purely formal character although it was formerly considered to be meaningful to ask whether Euclid's geometry was "true."

Theorems are most frequently referred to with a mathematical connotation. *See* LOGIC.

[P.W.B.]

the nucleus. The topic also includes the theory of nuclear reactions which is the basis of the technology of nuclear reactors. In an effort to understand the origin of nuclear forces, theoretical physicists have investigated the production and properties of mesons and the so-called strange particles. Radioactive decay, and particularly beta decay, is another branch of nuclear physics involving weak rather than strong nuclear forces. High-energy nuclear physics aims at understanding the properties of particles—nucleons as well as unstable particles of various kinds. *See* ELEMENTARY PARTICLE; MATHEMATICAL PHYSICS; NUCLEAR PHYSICS. [H.A.BE.]

Theory, physical

A physical theory usually involves the attempt to explain a certain class of physical phenomena by deducing them as necessary consequences of other phenomena regarded as more primitive and less in need of explanation. These more primitive phenomena may at the time the theory is formulated be undiscovered, so that part of the proof of the correctness of the theory consists in demonstrating the existence of the unknown assumed primitive phenomena. A classic example is the kinetic theory of gases in which the pressure of a gas is explained as arising from the kinetic reactions of colliding molecules, the reality of which was only established later by the discovery of phenomena such as the [illegible]

[illegible] in suggesting places to look for presently unknown new phenomena. [P.W.B.]

Theralite

A dark-colored, phaneritic (visibly crystalline) rock composed chiefly of pyroxene with smaller amounts of calcic plagioclase and nepheline. In a sense it is a nepheline-rich gabbro with abundant dark-colored (mafic) minerals.

Plagioclase is usually labradorite, and the pyroxene is titaniferous augite. Biotite mica, hornblende or barkevikite, olivine, and sodalite may all be present in minor amounts. Nepheline is an essential mineral; and as it diminishes somewhat in quantity, the rock may pass into essexite. A closely related rock is teschenite in which analcite proxies for nepheline. Accessory minerals include apatite, magnetite, and ilmenite.

[illegible]

[illegible] (molten rock) rich in alkalies and potential mafic minerals. It is associated with various types of gabbro, peridotite, and feldspathoidal rocks. It occurs in the Lugar sill in Ayrshire, Scotland, associated with considerable teschenite, picrite, and peridotite. *See* GABBRO; IGNEOUS ROCKS. [C.A.CA.]

Therapsida

The majority of advanced mammal-like reptiles are grouped into the order Therapsida (subclass Synapsida). Members of this group first appeared in mid-Permian times and persisted until the end of the Triassic. The group is highly ramified and includes several diverse lines of carnivores and herbivores. Most were terrestrial, but one group was aquatic. Therapsids are best known from deposits in South Africa but were world wide in distribution, with representation in the records of all continents except Antarctica.

The various evolving lines became increasingly mammalian through time. The most advanced types seem to have invaded uplands of the continental interiors. Among the therapsids appear to be the ancestors of mammals, which were derived either directly or through a synapsid group, Ictidosauria. *See* ICTIDOSAURIA; SYNAPSIDA. [E.C.O.]

Theria

One of the four subclasses of the class Mammalia, including all living mammals except the monotremes. The Theria were by far the most successful of the several mammalian stocks that arose from the mammal-like reptiles in the Triassic. The subclass is divided into three infraclasses: Pantotheria (no living survivors), Metatheria (marsupials), and Eutheria (placentals). These were not strictly contemporaneous; the Pantotheria arose directly from mammal-like reptiles, and the Metatheria and Eutheria in turn arose from pantotheres during the Cretaceous, many millions of years later. Therian mammals are characterized by the distinctive structural history of the molar teeth. The fossil record shows that all the extremely varied therian molar types were derived from a common tribosphenic type in which three main cusps, arranged in a triangle on the upper molar, are opposed to a reversed triangle and basinlike heel on the lower molar. *See* MAMMALIA. [D.D.D.]

Thermal expansion

Solids, liquids, and gases all exhibit dimensional changes for changes in temperature while pressure is held constant. The molecular mechanisms at work and the methods of data presentation are quite different for the three cases and are therefore discussed separately in this article.

Expansion of solids. The temperature coefficient of linear expansion α_l is defined by

$$\alpha_l = \frac{1}{l}\left(\frac{\partial l}{\partial t}\right)_{p=\text{const}}$$

where l is the length of the specimen, t is the temperature, and p is the pressure. For each solid there is a Debye characteristic temperature Θ, below which α_l is strongly dependent upon temperature and above which α_l is practically constant

theory of relativity has hardly anything in common with general relativity theory except the name, and in contrast to the general theory is established beyond doubt by an enormous number of experiments.) Newton's classical mechanics again is a special case of the mechanics of special relativity. *See* RELATIVISTIC ELECTRODYNAMICS; RELATIVISTIC MECHANICS.

Special relativity and quantum mechanics are examples of the development of physical theory. Neither of them has made classical mechanics wrong or obsolete, but they have extended classical mechanics into domains which were outside the experience of man until 1900. When a physical law is discovered, it can be expected to hold as long as the general conditions are not radically changed from those holding in the experiments from which the law was originally derived; for example, classical mechanics holds for objects of not too small size moving with moderate velocities. In a completely new area (for example, where there is very small size or very high speed) it cannot be expected a priori that the same laws will continue to hold; but if the laws do change under the new conditions, this does not invalidate the old laws in the domain for which they were originally formulated, except for minor corrections.

The most general theory of motion now known is quantum field theory, which combines both quantum mechanics and relativity theory and at the same time embodies the observed fact that particles can be created and annihilated. This theory may thus be called a "unified field theory." Attempts have also been made toward other unification, in particular to unify the theories of two types of forces, gravitational and electromagnetic; this is commonly called "unified field theory" in the literature. These attempts have so far not been very successful, moreover, they leave out quantum theory as well as the nuclear forces, both strong and weak. *See* QUANTUM FIELD THEORY; UNIFIED FIELD THEORIES.

Type of phenomena. The most customary classification of theoretical physics is according to the type of phenomena described. The following are the main fields under this heading:

1. Mechanics is the theory of motion of bodies under given forces. It is normally understood to involve classical mechanics only, and includes particle mechanics and mechanics of rigid bodies (*see* MECHANICS). In particle mechanics, celestial mechanics is an important subdivision; this includes planetary motion, the motion of artificial satellites, and the complicated motions resulting when three bodies interact (the classical three-body problem). Rigid-body mechanics includes the complicated theory of gyroscopic motion with and without external fields of force. *See* CELESTIAL MECHANICS; GYROSCOPE; RIGID-BODY DYNAMICS.

2. Continuum mechanics is the theory of motion of bodies taking into account their internal properties. One branch of this is the theory of elasticity which is basic for structural engineering design. Another branch is hydro- and aerodynamics. Here a number of problems can be solved approximately by potential theory, but most of modern aerodynamics requires a more physical approach. Knowledge of the physical properties such as those given by the equation of state of a gas is essential; these properties can be explained only on a molecular scale. Acoustics is a classical branch of continuum mechanics. A combination of aerodynamics and electrodynamics is required for the modern field of magnetohydrodynamics. *See* ACOUSTICS; AERODYNAMICS; ELASTICITY; FIELD THEORY, CLASSICAL; HYDRODYNAMICS; MAGNETOHYDRODYNAMICS; POTENTIALS (PHYSICS).

3. Heat presents a problem that can be treated on a phenomenological level by thermodynamics, which is the basis of heat engineering as well as of the theory of chemical equilibrium. On the molecular level, heat is described by statistical mechanics, which may be considered the physical foundation of thermodynamics. Beyond this, statistical mechanics permits the calculation of the properties of bulk substances (gases, liquids, and solids) in terms of atomic properties. *See* HEAT; STATISTICAL MECHANICS; THERMODYNAMIC PRINCIPLES.

4. Electrodynamics is well understood. Subdivisions are electrostatics; the theory of stationary currents (the basis of electrical generating machinery); the theory of oscillating electrical circuits (the basis of the technology of ordinary radio); the theory of electromagnetic waves, including their propagation in air as well as in wave guides and similar devices (the basis of radar); and finally the electromagnetic theory of light. *See* ELECTRODYNAMICS.

5. Optics is customarily treated as a special field although, strictly speaking, it is a branch of electrodynamics. Geometrical optics and the theory of diffraction phenomena are two of the principal topics. Emission and absorption of light can be understood only on the basis of atomic physics. The same is true of dispersion, that is, the behavior of the refractive index as a function of frequency. *See* OPTICS.

6. Atomic physics includes the theory of the structure of the atom, the motion of the electrons in the atom, the periodic system; the energy levels and spectral lines of atoms and molecules; the behavior of atoms and molecules in external fields, and collisions of atoms with each other, with electrons, and with other particles. Atomic physics is the basis of the calculation of properties of matter in bulk and of the emission and absorption of light. Related is the theory of molecular structure which is the basis of theoretical chemistry. Collisions between molecules explain the rate of chemical reactions. *See* ATOMIC PHYSICS; MOLECULAR PHYSICS.

7. Nuclear and particle physics includes the theory of nuclear forces and of the structure of atomic nuclei. A complete theory would predict all energy levels of any nucleus, and thus the electromagnetic radiations which can be emitted by

the nucleus. The topic also includes the theory of nuclear reactions which is the basis of the technology of nuclear reactors. In an effort to understand the origin of nuclear forces, theoretical physicists have investigated the production and properties of mesons and the so-called strange particles. Radioactive decay, and particularly beta decay, is another branch of nuclear physics involving weak rather than strong nuclear forces. High-energy nuclear physics aims at understanding the properties of particles—nucleons as well as unstable particles of various kinds. *See* ELEMENTARY PARTICLE; MATHEMATICAL PHYSICS; NUCLEAR PHYSICS. [H.A.BE.]

Theory, physical

A physical theory usually involves the attempt to explain a certain class of physical phenomena by deducing them as necessary consequences of other phenomena regarded as more primitive and less in need of explanation. These more primitive phenomena may at the time the theory is formulated be undiscovered, so that part of the proof of the correctness of the theory consists in demonstrating the existence of the unknown assumed primitive phenomena. A classic example is the kinetic theory of gases, in which the pressure of a gas is explained as arising from the kinetic reactions of colliding molecules, the reality of which was only established later by the discovery of phenomena such as the Brownian fluctuations. The value of a theory depends both on the success with which it coordinates [illegible]

Theralite

A dark-colored, phaneritic (visibly crystalline) rock composed chiefly of pyroxene with smaller amounts of calcic plagioclase and nepheline. In a sense it is a nepheline-rich gabbro with abundant dark-colored (mafic) minerals.

Plagioclase is usually labradorite, and the pyroxene is titaniferous augite. Biotite mica, hornblende or barkevikite, olivine, and sodalite may all be present in minor amounts. Nepheline is an essential mineral; and as it diminishes somewhat in quantity, the rock may pass into essexite. A closely related rock is teschenite in which analcite proxies for nepheline. Accessory minerals include apatite, magnetite, and ilmenite.

[illegible] magma (molten rock) rich in alkalies and potential mafic minerals. It is associated with various [illegible] oidal [illegible] Scot- [illegible] pic- [illegible] and peridotite. *See* GABBRO; IGNEOUS ROCKS [C A CA.]

Therapsida

The majority of advanced mammal-like reptiles are grouped into the order Therapsida (subclass Synapsida). Members of this group first appeared in mid-Permian times and persisted until the end of the Triassic. The group is highly ramified and includes several diverse lines of carnivores and herbivores. Most were terrestrial, but one group was aquatic. Therapsids are best known from deposits in South Africa but were world wide in distribution, with representation in the records of all continents except Antarctica.

The various evolving lines became increasingly mammalian through time. The most advanced types seem to have invaded uplands of the continental interiors. Among the therapsids appear to be the ancestors of mammals, which were derived either directly or through a synapsid group, Ictidosauria. *See* ICTIDOSAURIA, SYNAPSIDA. [E C O.]

Theria

One of the four subclasses of the class Mammalia, including all living mammals except the monotremes. The Theria were by far the most successful of the several mammalian stocks that arose from the mammal-like reptiles in the Triassic. The subclass is divided into three infraclasses: Pantotheria (no living survivors), Metatheria (marsupials), and Eutheria (placentals). These were not strictly contemporaneous; the Pantotheria arose directly from mammal-like reptiles, and the Metatheria and Eutheria in turn arose from pantotheres during the Cretaceous, many millions of years later. Therian mammals are characterized by the distinctive structural history of the molar teeth. The fossil record shows that all the extremely varied therian molar types were derived from a common tribosphenic type in which three main cusps, arranged in a triangle on the upper molar, are opposed to a reversed triangle and basinlike heel on the lower molar. *See* MAMMALIA [D D D.]

Thermal expansion

Solids, liquids, and gases all exhibit dimensional changes for changes in temperature while pressure is held constant. The molecular mechanisms at work and the methods of data presentation are quite different for the three cases and are therefore discussed separately in this article.

Expansion of solids. The temperature coefficient of linear expansion α_l is defined by

$$\alpha_l = \frac{1}{l}\left(\frac{\partial l}{\partial t}\right)_{p=\text{const}}$$

where l is the length of the specimen, t is the temperature, and p is the pressure. For each solid there is a Debye characteristic temperature Θ, below which α_l is strongly dependent upon temperature and above which α_l is practically constant

Many common substances are near or above Θ at room temperature and follow the approximate equation

$$l = l_0(1 + \alpha_l t)$$

where l_0 is the length at 0°C and t is the temperature in °C. The total change in length from absolute zero to the melting point ranges around 2% for most substances. Typical room temperature values of α_l are as follows:

Substance	*Coefficient of linear expansion per °C* (× 10^{-6})
Aluminum, commercial	24
Copper	17
Diamond	1
Glass, commercial	11
Glass, pyrex	3
Granite	8.3
Ice	50
Iron	12
Invar alloy	0.9
Quartz, crystalline	5
Quartz, fused	0.5
Oak, along fiber	5
Oak, across fiber	54
Rubber, hard	80

Linear, harmonic vibration of the atoms in a solid cannot account for changes in volume, hence this must result from nonlinearity of the thermally excited vibration The theory of E. Gruneisen takes this into account and shows the coefficient of expansion to be proportional to the constant-volume specific heat of the solid. At low temperatures (small amplitude vibration), the coefficient of expansion approaches zero.

Pure crystals may have different values of α_l along different axes, but substances like structural steel have many crystals randomly oriented and are almost free from this effect. At certain temperatures crystalline substances may change in lattice arrangement, and a sudden change of volume will occur at constant temperature, making α_l momentarily infinity. *See* LATTICE VIBRATIONS, SPECIFIC HEAT OF SOLIDS; THERMOCOUPLE [J.D L.]

Expansion of gases. So-called perfect gases follow the equation

$$pv/T = R/\text{molecular weight}$$

where p is absolute pressure, v is specific volume, T is absolute temperature, and R is a constant The magnitude of R, the so-called gas constant, is 1544 ft-lb/(°Rankine) (pound-mole) in the English system, or 8.3144×10^7 ergs/(°Kelvin) (gram-mole) in the metric system (*see* GAS CONSTANT). Real gases often follow this equation closely; for example, the following list shows values of R at atmospheric pressure and 0°C:

Gas	*R*
Air	1545
Hydrogen	1546
Nitrogen	1543
Oxygen	1544
Methane	1539

The coefficient of cubic expansion α_v is defined by

$$\alpha_v = \frac{1}{v}\left(\frac{\partial v}{\partial t}\right)_{p=\text{const}}$$

and for a perfect gas this is found to be $1/T$.

The behavior of real gases is largely accounted for by van der Waal's equation

$$p = \frac{RT}{v - b} - \frac{a}{v^2}$$

where a and b are constant for a given gas. When the specific volume is large, the effects of these constants are unimportant, and the real gas behaves as a perfect gas. In the regions where a and b have a dominant effect it is usually found desirable to use experimentally determined graphs or charts of properties. *See* GAS; KINETIC THEORY OF MATTER.

Expansion of liquids. For liquids, α_v is somewhat a function of pressure but is largely determined by temperature. Though α_v may often be taken as constant over a sizeable range of temperature (as in the liquid expansion thermometer), generally some variation must be accounted for. For example, water contracts with temperature rise from 0°C to 4°C, above which it expands at an increasing rate, as shown by the following data. which were taken at atmospheric pressure:

t, °C	*Volume expansion, ml/g*
-10	1.00186
0	1.00013
4	1.00000
10	1.00027
100	1.007

One approach to this variation is to evaluate the constants α, β, and γ in the equation

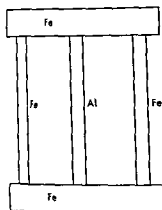
$$v = v_0(1 + \alpha t + \beta t^2 + \gamma t^3)$$

where v_0 is the volume at 0°C. and v is the volume at temperature t Typical values of the coefficients appear in the accompanying table.

Coefficients of volume expansion of gases

Liquid	α	β	γ
Ethyl alcohol (99.3% by volume)	1.012	2.20	
500 atmospheres	0.866		
3000 atmospheres	0.524		
Carbon tetrachloride	1.184	0.899	1.351
Mercury	1.182	0.0078	
Petroleum	0.8994	1.396	
Water	−0.06427	8.5053	−6.7900

Thermal stresses. When a homogeneous body is subject to constant boundary loads and is raised uniformly in temperature, the stress pattern in it will not change unless its elastic properties change. In per[illegible] made [illegible] of expansion, (2) changes of [illegible] are restrained, or (3) temperature distribution is not uniform. A simple example of the first case is shown in the figure, where the aluminum bar, if



Thermal stresses would arise in the above body if it were subjected to heat.

heated, would tend to expand faster than the iron bars, thereby putting the iron in tension and the aluminum in compression. Considering the alumi- [illegible]

[illegible] and if this one alone were heated, again the other bars would be in tension and the center one in compression. More complex stress patterns may arise in continuous bodies; for example, if the bars were joined along the sides rather than at the ends, shear stresses would arise in the seams. In iron, 300 lb/in² tensile stress would produce the same elongation as would a temperature rise of 1°C.

Since one source of temperature variation is the gradient necessary for heat transfer, thermal conductivity and heat capacity may both play a role in determining the stress pattern. See THERMAL STRESS; see also CONDUCTION (HEAT); HEAT CAPACITY; THERMOMETER. [R A BU.]

Thermal neutrons

Neutrons whose energy distribution is identical with, or similar to, the thermal distribution of the material in which they are found. Other definitions of thermal neutrons are (1) that part of the neutron spectrum which has a Maxwell-Boltzmann energy distribution corresponding to the environmental temperature, that is, the maximum Maxwell Boltzmann distribution which can be subtracted from the total spectrum to yield a nonnegative remainder, (2) all neutrons with energies less than 0.4 ev, or the energy at which the cadmium absorption cross section becomes small. The average energy of thermal neutrons at room temperature is about 0.025 ev.

In some cases, the energy distribution of the neutrons is not in thermal equilibrium with the environment, but they are still called thermal distributions of neutrons. Most neutrons are produced at energies above thermal, and the absorption of neutrons interrupts the process of slowing down so that the low-energy neutron spectrum is "hardened" with a greater number of neutrons of higher energy than the Maxwellian distribution of energies predicts, and a "tail" of neutrons not yet thermalized. The actual distribution can be fitted to one in which the spectrum is represented as a Maxwellian distribution of higher temperature than the environment, with the tail superposed, this fitting procedure is often useful in predicting reaction rates. In this case, the altered Maxwellian is used to define a thermal neutron spectrum.

In a system of high leakage, with mean free path of neutrons constant, the spectrum is "colder" than that of the environment, if the same fitting procedure as described above is used. See BOLTZMANN STATISTICS; FISSION, NUCLEAR; KINETIC THEORY OF MATTER, REACTOR PHYSICS [B.I S.]

Thermal spring

A spring with water temperature substantially [illegible]

[illegible] of springs is [illegible] enheit of the mean annual temperature of the atmosphere. Thus, waters of thermal springs range in temperature from as low as 60°F, in an area where

[illegible]

thermal springs are the source of the water and the source of the heat. The water may be ordinary ground water that percolates slowly downward, is heated by the earth's normal thermal gradient (the temperature of the earth normally increases about 1°F for each 50–100 ft of depth), and then returns to the surface without losing all the added heat. The water of thermal springs may be in part juvenile, a product of the crystallization or recrystallization of rock at depth. Juvenile water is virtually certain to become mixed with connate or meteoric water on its way to the surface, and there are no thermal springs whose water can be demonstrated to be wholly juvenile.

Investigations of Warm Springs, Georgia, and of other thermal waters in the eastern United States indicate that the water entered the aquifer by normal recharge from precipitation, percolated deep into the earth by reason of the geologic structure, and there received its heat before returning to the surface. On the other hand, the springs in Yellowstone Park in Wyoming, Steamboat Springs in Nevada, and many other localities in the western United States may derive part of their water and

much of their heat from bodies of superheated rocks, perhaps in the last stages of cooling from the molten state. Many of these latter described springs discharge water that is near the boiling point.

The total average discharge of all the thermal springs in the United States has been estimated at 700,000,000 gal/day, which is about the same as the discharge of Silver Springs in Florida The contribution of thermal springs is, therefore, only a small fraction of the water discharged by all springs. [A.N.S.]

Bibliography: G. A. Waring, *Springs of California*, USGS Water Supply Paper 338, 1915.

Thermal stress

Mechanical stress induced in a body when some or all of its parts are not free to expand or contract because of changes in temperature In most continuous bodies, thermal expansion or contraction cannot occur freely in all directions, and so stresses are produced In addition, the external restraints on the body that prevent thermal expansion or contraction also produce stresses in the body. These stresses, known as thermal stresses, can be either a normal stress or a bending stress, or a combination of both

Structures subject to stress. Problems of thermal stress arise in the design of steam and gas turbines, diesel engines, jet engines, and rocket motors. The design of air-frames presents even more severe problems because of the temperature encountered These problems are further complicated because of transient heating, nonsimple geometry, and properties of material at high temperature Thermal stress is also a factor in the design of atomic power plants.

The thermal stress problem is no longer restricted to the classical case of finding the elastic thermal stresses for a given temperature distribution in a structure with no buckling. It touches on all phases of the structural design, including the temperature distribution, elastic and inelastic deformation, allowable stresses for various materials and loading conditions, and the buckling, deflection, stiffness, fatigue, shock, and aeroelastic effects of elevated temperatures

The method of attack for these problems is to set up the complete problem and then simplify it by making assumptions based on the physical situation or on experimental data. An analytical solution is obtained for the approximate problem, which demonstrates the basic parameters in the problem and allows charts to be constructed, showing how the stresses vary with these parameters. Refinements in the solution are then made by investigating the simplifying assumptions and obtaining correction factors. In most cases, this procedure yields results with sufficient accuracy for use in the design of the structure.

Procedures at elevated temperatures. To design for thermal stress in aircraft and missiles, it is necessary to know the expected mission profiles or how the flight histories of speed, altitud and angle of attack vary with time. From the predicted mission profiles for various vehicles, th worst combinations of applied, thermal, and allov able stresses are selected to be used in the desig

At elevated temperatures, the designer is face with two factors which tend to increase the weig of the structure. First, the strength and moduli o most materials are much less than at room ten perature Second, the thermal stresses may add t the applied stresses and may further reduce th torsional stiffness of a structure. To obtain th lightest possible structure, both of these problem must be considered. For flights of short duratio insulation and radiation cooling may help keep th material temperature low and reduce the effects o thermal stresses. For longer flights, insulation plu cooling can minimize the temperature.

A mechanical design which allows sufficient de flection for thermal expansion can relieve part o thermal stress. To maintain structural stiffness i not always simple. Expansion joints may be diffi cult to design Corrugated webs, ribs, and cli attachments may relieve some thermal stresses an still maintain stiffness. However, heavy spar cap or thick skin with integral stiffeners have high thermal stresses. One possible way to relieve ther mal stress is to let the inelastic portion of th stress-strain curve provide the deflection to absorl the thermal expansion. This procedure involve the concept of strain design rather than stress design, but it is possible in many cases to design for the applied loads without regard to thermal stresses, and then add the thermal strains without obtaining appreciable permanent set In other cases, a design which allows for some permanent set may be feasible However, if the structure is exposed to numerous temperature cycles, thermal fatigue must be considered for such designs using inelastic stresses

Allowable stresses. Combined thermal and applied stresses at elevated temperatures may be either in the elastic state or in the inelastic range for structural materials. If they are in the inelastic range, mechanical properties of the structural material must be determined for various temperatures To design the structure to support the applied and thermal stresses, it is also necessary to know the allowable stresses of the material under various loading and temperature conditions. A honeycomb-sandwich-structure design can increase the strength of a material at elevated temperature for a given weight

Creep. For most materials, creep rate increases substantially with temperature and stress. Materials used in aircraft and missile structures creep at elevated temperatures so that the structure may have large deformations if the load is applied for long periods. To keep these deformations within permissible limits, it may be necessary to reduce the applied stresses on the structure.

Thermal fatigue. At room temperature, fatigue limits the number of stress cycles a material can

withstand before it ruptures. Thermal fatigue is caused by both the stress and the temperature cycles and can result in rupture and deformation arising from creep. When a structure is restrained, it is possible for sufficiently large temperatures to produce thermal stresses in the inelastic region of the stress-strain curve that exceed yield stress and result in plastic flow or rupture.

Thermal shock. If a body is subjected to a steep transient temperature gradient so that large thermal stresses are induced, thermal shock is produced. Such shock arises when a body at one uniform temperature is suddenly accelerated to or decelerated from high supersonic or hypersonic speeds. A sudden change of angle of attack produces the same effect, even when constant speed is maintained. Thermal stress can be either a normal stress or a bending stress, or a combination of both. In general, if the temperature distributions are symmetrical with respect to axes of symmetry in the symmetrical section, the thermal stress induced will not involve bending stresses. However, when the structure temperature is nonuniform bending stresses may be induced by an unsymmetrical temperature distribution, by an unsymmetrical section, by different materials in the structure, or by variation of physical properties with respect to temperature. Brittle and ductile materials react differently to such thermal stresses. Because the thermal stress arises from the strain due to temperature expansion, *brittle materials* which can endure little strain before rupture, *may* fail under the thermal shock. See AEROELASTICITY; SPACECRAFT STRUCTURE [S.Y.C.]

Bibliography: R. L. Bisplinghoff, Some structural and aeroelastic considerations of high-speed flight, *J. Aeronaut. Sci.*, 23(4):289–321, 1956; B. E. Gatewood, *Thermal Stresses*, 1957; N. J. Hoff, Structure problems of future aircraft, *J. Roy. Aeronaut Soc.*, 55:678, 1951.

Thermionic emission

The emission of electrons into vacuum by a heated electronic conductor. In its broadest meaning, thermionic emission includes the emission of ions, but since this process is quite different from that normally understood by the term, it will not be discussed h[illegible]

[illegible] conductors are thermionic emitters, only a few materials satisfy the requirements set by practical applications. Of the metals, tungsten is an important practical thermionic emitter; in most electron tubes, however, the oxide-coated cathode is used to great advantage. For a detailed discussion of practical thermionic emitters see VACUUM TUBE.

Richardson equation. The thermionic emission of a certain material may be measured by using the material as the cathode in a vacuum tube and collecting the emitted electrons on a positive anode. If the anode is sufficiently positive relative to the cathode, space-charge (a concentration of electrons near the cathode) may be avoided and *all* electrons emitted are collected; one then measures the saturation thermionic current. Actually, the emission current increases slightly with increasing field strength at the cathode, and in order to obtain the true saturation current one should extrapolate to zero applied field (*see* SCHOTTKY EFFECT). The emission current density J increases rapidly with increasing temperature; this is illustrated by the following approximate values for tungsten:

T (°K)	1000	2000	2500	3000
J (amp/cm²)	10^{-15}	10^{-3}	0.3	15

The temperature dependence of J is given by the Richardson (or Dushman-Richardson) equation

$$J = AT^2e^{-(\phi/kT)}$$

Here A is a constant; k is Boltzmann's constant ($= 1.38 \times 10^{-23}$ joule/degree) and ϕ is the work function of the emitter. The work function has the dimensions of energy and is of the order of a few electron volts for thermionic emitters. For a table of values for metals, *see* WORK FUNCTION (ELECTRONIC). The temperature dependence of J is essentially determined by the exponential factor, since its temperature dependence predominates strongly over that of the factor T^2. Both A and ϕ may be obtained experimentally by plotting the logarithm of J/T^2 versus $1/T$, as illustrated for tungsten in Fig. 1. The Richardson formula can be derived for metals and semiconductors on the basis of relatively simple physical models.

Metals. According to quantum theory the electrons in a free atom occupy a set of discrete en-

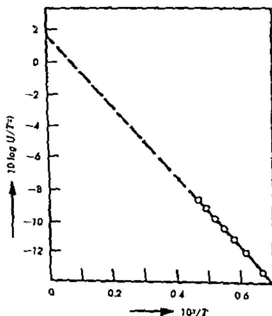


Fig. 1. Richardson plot for tungsten (After G. Herrmann and S. Wagener, The Oxide-Coated Cathode, vol. 2, Chapman and Hall, 1951)

ergy levels. When atoms are brought together to form a solid, these energy levels broaden into energy bands; the broadening is a result of the perturbing fields produced by neighboring atoms on the electrons and is most pronounced for the outer or valence electrons (*see* BAND THEORY OF SOLIDS). *In a metal,* the perturbing influence on the valence electrons is so strong that they can no longer be associated with particular atoms but must be considered as moving freely throughout the crystal. These so-called free, or conduction, electrons are responsible for the high electrical and thermal conductivity of metals and also for the thermionic emission. *See* FREE-ELECTRON THEORY OF METALS.

The free electrons may be assumed to move in an approximately constant potential as indicated in Fig. 2. The bottom of the box corresponds to the energy of a conduction electron at rest in the metal; the "vacuum" level represents the energy of an electron at rest *in free space. According to* quantum mechanics, the electrons in this model can assume only particular states of motion which correspond to a set of very closely spaced energy levels. The probability for a given state to be occupied depends on the energy E of the state and on the absolute temperature T in accordance with the so-called Fermi-Dirac distribution function

$$F(E) = \frac{1}{1 + \exp\left[(E - E_F)/kT\right]}$$

The quantity E_F is called the Fermi energy, it is determined by the number of electrons per unit volume in the metal and is of the order of a few electron volts. Since kT at room temperature ($T = 300°K$) is only about 0 025 electron volt, $E_F \gg kT$ for all temperatures below the melting point of metals. Note that for $T = 0$, $F(E) = 1$ for $E < E_F$, and $F(E) = 0$ for $E > E_F$ Hence, at absolute zero all energy levels up to E_F are occupied by electrons, whereas those above E_F are empty For temperatures different from zero, some electrons have energies larger than E_F and the thermionic emission is due to those electrons in the "tail" of the Fermi distrib. [illegible] the vacuum le[illegible] $F(E) = 0.5$; [illegible] to those states for which the probability of being occupied is equal to 0.5.

When these ideas are put in a quantitative form, one arrives at the Richardson equation with the specific value $A = 120$ amp/cm² (if one takes into account reflection of electrons against the surface potential barrier, the theoretical value of A is < 120 amp/cm²).

Experiments by M. N. Nichols in 1958 and by G. F. Smith in 1954 on single crystals of tungsten have shown that experimental values for A and ϕ depend on the crystallographic plane from which the emission is measured; values for A (in amp/cm²) and ϕ (in electron volts) for two crystallographic directions are given in the table. For polycrystalline metals, the experimental values for A and ϕ are thus average values for the particular specimen.

Direction	Nichols		Smith	
	A	ϕ	A	ϕ
(111)	35	4 39	52	4 38
(100)	117	4 56	105	4 52

Semiconductors. For semiconductors, the thermionic emission is also due to the escape of electrons which have energies above the vacuum level. The theory leads to the Richardson formula, as it does for metals. The work function measures again the difference between the Fermi level of the semiconductor and the vacuum level. [A.J DE.]

Bibliography: S. Fluegge (ed.), *Handbuch der Physik*, vol. 21, 1956

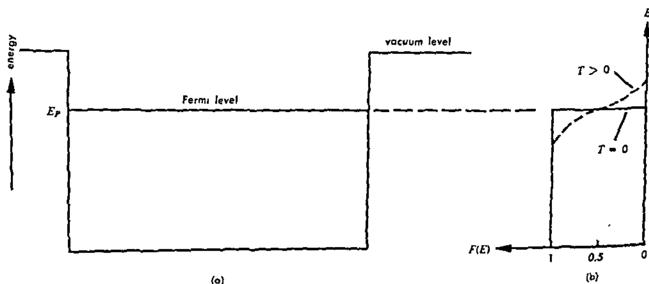


Fig. 2. *(a) Occupation of electron states between the* bottom of the conduction band and the Fermi level of a metal is indicated for $T = 0$ by the shaded area. *(b) Fermi distribution function is represented schematically for $T = 0$ and for $T > 0$.*

Thermionic tube

An electron tube that relies upon thermally emitted electrons from a heated cathode for tube current.

Thermionic emission of electrons means emission by heat. In practical form an electrode, called the cathode, is heated until it emits electrons. The cathode may be either a directly heated filament or an indirectly heated surface. With a filamentary cathode, heating current is passed directly through the wire, which either emits electrons directly or is

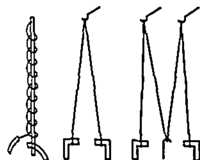


Fig. 1 Typical filament structures.

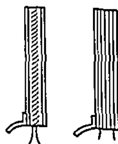


Fig. 2 Indirectly heated cathodes.

covered with a material that readily emits electrons. Some typical filament structures are shown in Fig. 1. See THERMIONIC EMISSION.

Indirectly heated cathodes commonly have the forms shown in Fig. 2. In these cathodes a filament, located within the cathode electrode, carries the heating current. This form is most commonly used when the cathodes are coated with barium and strontium oxide. An indirectly heated cathode is often referred to as an equipotential cathode, indicating that there is no voltage drop along the emission surface as with dire[illegible]

[illegible] called cold-cathode tubes, but the current is generally not as stable as with thermionic cathodes. See ELECTRON TUBE; VACUUM TUBE [K.R.S.]

Thermistor

A resistive circuit component, having a high negative temperature coefficient of resistance (as temperature increases, resistance decreases). See RESISTOR, ELECTRICAL.

The typical thermistor is a stable, compact, and rugged two-terminal ceramiclike semiconductor bead, rod, or disk. Thermistors are made of various mixtures of oxides of manganese, nickel, cobalt, copper, uranium, iron, zinc, titanium, and magnesium. The temperature coefficient is determined by the proportions of oxides in the mixture.

In bead thermistors, the oxide mixture is applied to two parallel platinum wires (0.001–0.005 in. in diameter) as a viscous droplet (0.006–0.060 in. in diameter). Upon firing, the ceramic bead cements the wires which become the leads. The wire diameter, bead size, mixture, and wire spacing determine the device characteristics. Bead thermistors have little mass and a short time constant.

In disk and washer thermistors, an oxide-binder mixture is pressed and sintered. The disk type is 0.200–0.600 in. in diameter and 0.040–0.500 in. thick. The washer type may be as large as 0.750 in. in diameter and 0.500 in. thick. The major surfaces are coated with a conducting material and leads attached. The thin, large-diameter disks have low resistance, short time constant and high power dissipation. Thick, small diameter units have high resistance, long time constant and low power dissipation.

Rod-type thermistors are extruded as long, slim rods (0.250–2.0 in. long and 0.050–0.110 in. in diameter) of oxide-binder mix and sintered. The ends are coated with conducting paste and leads are wrapped on the coated area. The rod type has high resistance, long time constant and moderate power dissipation.

Characteristics. Thermistors are characterized by size, shape, temperature coefficient, and resistance. The dissipation constant is the power dissipated divided by the temperature rise above ambient. Power sensitivity is the watts dissipation required to reduce resistance 1%. Time constant is the time required for the device temperature to fall 63% toward ambient.

Temperature-resistance characteristic. One typical thermistor mix of oxides of manganese and nickel has a temperature coefficient of resistance of −4.4% per °C at 25°C. For comparison, the coefficient of copper is +0.39% per °C. From −100°C to +400°C, resistance will change 10,000,000:1 as shown by curve 1 of the illustration.

Another typical thermistor mix comprising oxides of manganese, nickel, and cobalt has a coefficient of −3.9% per °C at 25°C (see curve 2 of the illustration).

Voltage current characteristic. This characteristic of a thermistor is a function of its dissipation constant. With a small current flowing, voltage drop will follow Ohm's law. As power is dissipated, device temperature rises, resistance lowers, and voltage drop is correspondingly reduced. With further increase in current (and allowing thermal equilibrium to be reached at each step) temperature rises and resistance decreases. The voltage drop increases, but is less than it would have been had the resistance stayed constant. At some value

of current, the voltage reaches a maximum. At higher values of current, temperature rises but voltage drop decreases. In this region, the thermistor has negative resistance.

Typical applications of thermistors. Thermistors are versatile circuit elements and have many applications in measurement and control. Some typical, important applications are discussed below.

Temperature measurement. The thermistor's large temperature coefficient of resistance is ideal for temperature measurement. In this application, the power dissipated must be so small that it does not heat the device.

Precise temperature measurement is done with high-resistance thermistors in a resistance bridge. Sensitivity of 0.0005°C is readily attained. Lead resistance has no effect. Compensating leads and cold junctions are unnecessary. Bead thermistors are built into equipment at locations where temperature is to be measured (gear housings, bearings, cylinder heads, transformer cores) and a resistance bridge measures the temperatures at a remote station. See TEMPERATURE MEASUREMENT; THERMOMETER.

Temperature compensation. Many electrical and electronic components have positive temperature coefficients which are detrimental to the temperature stability of the circuit. A properly-selected thermistor in the circuit containing such a component will provide temperature compensation.

Flowmeter, vacuum gage, anemometer. A small voltage is applied and the current through the thermistor is measured. The amount of heat dissipated is a function of the degree of vacuum surrounding the device or the velocity of gas passing the device, and the dissipated power reaches a constant value. The measured current is calibrated in terms of vacuum or gas flow. See ANEMOMETE[R]; FLOW MEASUREM[ENT]; V...

Time delay. T
a result of cur[r]
ance decreases.
rate of decrease
circuit current can be used to introduce a fix[ed] time delay between relay operations or to prote[ct] equipment during start-up.

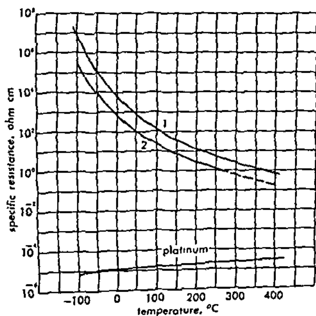
Power measurement, bolometer. The thermistor resistance versus power characteristic makes it useful power-measuring device. Microwave powe[r] is measured by a bead thermistor mounted in th[e] waveguide and biased so that bead impedanc[e] matches the cavity. When radio-frequency powe[r] is applied, the bead is heated by absorbed powe[r]. The bias current is reduced so that the thermisto[r] remains at the same operating temperature. Th[e] decrease in bias power is just equal to the radio frequency power absorbed. The thermistor also ca[n] be used to measure radiant power, such as infrare[d] or visible light. See ELECTRIC POWER MEASUREMENT.

Other applications. Thermistors are used as voltage regulators and volume limiters in communication circuits. A shunt voltage regulator is provided by shunting the circuit with a suitably chosen value of resistance in series with a thermistor. Networks of resistors and thermistors are used as compressors, expanders and limiters in transmission circuits. [F.H.B]

Bibliography: J. A. Becker, C. B. Green, G. L. Pearson, Properties and uses of thermistors—thermally sensitive resistors, *Trans AIEE*, 65:711–725, 1946; Victory Engineering Corp., *Thermistors Data Book*.

Thermoanalysis

A group of techniques for continuously measuring the effects attending chemical or physical changes caused by various processes that occur in a single or multicomponent system as the temperature is varied at a selected heating rate. A linearly increasing temperature cycle of 5–15°C per minute is commonly used, and the measurements are conducted in an environment of air or other static or dynamically controlled atmosphere at elevated, ambient, or reduced pressures. The resulting thermograms describe the system uniquely in terms of the heat effects, changes in weight, volume of gas evolved or absorbed, or one of many other characteristic changes in physicochemical properties and reactions that occur as functions of temperature. The principal thermoanalytical methods are thermogravimetry (TG) and differential thermal analysis (DTA). Instrumentation required for these complementary techniques includes the sample holder contained in a furnace equipped with a temperature programmer, and a means for converting the physicochemical property being measured into an electrical signal that can be measured as a function of the temperature, or of time, con-



Specific resistance versus temperature for manganese-nickel thermistor mix (curve 1) and manganese-nickel-cobalt mix (curve 2) compared with a metal.

ly with the sample, furnace, or reference rature.

s the combination of temperature ranges over n the relative changes in weight, heat content, ther physicochemical properties take place the rates at which they occur that uniquely esent the system Measurements of the temture ranges and rates provide experimental that can be used for the development of anical research and control procedures as well as other physicochemical investigative purposes. e intrinsic value of these techniques lies in the errelationship between the thermodynamic and emical kinetic parameters associated with the lative thermal stability of chemical and physical nds formed or disrupted. These bonds are rmed or disrupted when the heated substance ndergoes changes in state, adsorption, desorption r absorption, corrosion, addition reactions, or ecomposition.

Thermogravimetry. This method involves measuring the changes in weight of a substance as it is heated to elevated temperatures. The thermobalance required for this thermoanalytical technique consists of a precision balance and furnace that have been adapted for continuously measuring or [illegible]

chemical reactions that involve either a gain or a loss in weight may be studied by this method. Rates of reaction and energies of activation for vaporization, sublimation, and chemical reaction can be obtained in addition to changes in weight. Those characterized by a gain in weight include

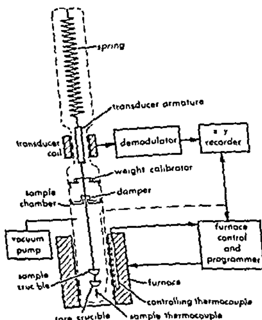


Fig. 1 Schematic diagram of an automatic recording vacuum thermobalance (American Instrument Co.)

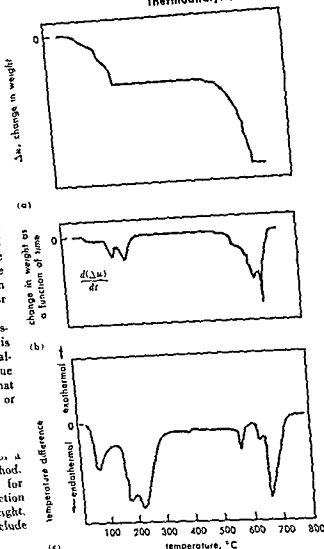


Fig. 2 Thermoanalytical curves for calcium nitrate tetrahydrate, $Ca(NO_3)_2 \cdot 4H_2O$. (a) Thermogravimetric analysis (b) Derivative thermogravimetry (c) Differential thermal analysis

adsorption and absorption of gases or vapors, corrosion in inorganic or organic oxidizing atmospheres, and direct combination with gaseous reactants A loss in weight results from the desorption of gases or vapors, vaporization of liquids, sublimation of solids, dehydration, desolvation, the gas-evolving decomposition of inorganic or organic substances, and the degradation of polymeric complexes.

Applications of thermogravimetric analysis (TGA) have been concerned primarily with establishing the temperature ranges required for the proper drying or pyrolysis of the inorganic, organic, and metalloorganic compounds used as stoichiometric precipitates in gravimetric analysis. Mixtures of hydrated materials are amenable to this type of analysis because of the thermodynamic equilibrium conditions associated with the stepwise stages of dehydration that occur as the temperature is raised. More important, this analytical method provides data that often can be used to

of current, the voltage reaches a maximum. At higher values of current, temperature rises but voltage drop decreases. In this region, the thermistor has negative resistance.

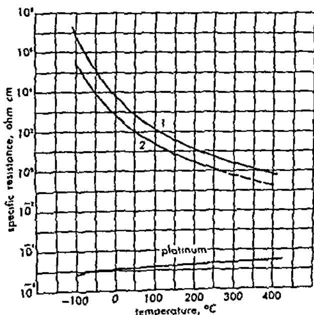
Typical applications of thermistors. Thermistors are versatile circuit elements and have many applications in measurement and control. Some typical, important applications are discussed below.

Temperature measurement. The thermistor's large temperature coefficient of resistance is ideal for temperature measurement. In this application, the power dissipated must be so small that it does not heat the device.

Precise temperature measurement is done with high-resistance thermistors in a resistance bridge. Sensitivity of 0.0005°C is readily attained. Lead resistance has no effect. Compensating leads and cold junctions are unnecessary. Bead thermistors are built into equipment at locations where temperature is to be measured (gear housings, bearings, cylinder heads, transformer cores) and a resistance bridge measures the temperatures at a remote station. See TEMPERATURE MEASUREMENT; THERMOMETER.

Temperature compensation. Many electrical and electronic components have positive temperature coefficients which are detrimental to the temperature stability of the circuit. A properly-selected thermistor in the circuit containing such a component will provide temperature compensation.

Flowmeter, vacuum gage, anemometer. A small voltage is applied and the current through the thermistor is measured. The amount of heat dissipated is a function of the degree of vacuum surrounding the device or the velocity of gas passing the device, and the dissipated power reaches a constant value. The measured current is calibrated



Specific resistance versus temperature for manganese-nickel thermistor mix (curve 1) and manganese-nickel-cobalt mix (curve 2) compared with a metal.

in terms of vacuum or gas flow. See ANEMOMETI FLOW MEASUREMENT; VACUUM MEASUREMENT.

Time delay. When a thermistor is self-heated a result of current passing through it, its resi ance decreases. Due to the thermal mass, the tir rate of decrease is fixed. The delayed build-up circuit current can be used to introduce a fix time delay between relay operations or to prote equipment during start-up.

Power measurement, bolometer. The thermistoı resistance versus power characteristic makes it useful power-measuring device. Microwave pow is measured by a bead thermistor mounted in tl waveguide and biased so that bead impedan matches the cavity. When radio-frequency pow is applied, the bead is heated by absorbed powe The bias current is reduced so that the thermisto remains at the same operating temperature. Th decrease in bias power is just equal to the radio frequency power absorbed. The thermistor also ca be used to measure radiant power, such as infrare or visible light. See ELECTRIC POWER MEASURE MENT.

Other applications. Thermistors are used as voltage regulators and volume limiters in communication circuits. A shunt voltage regulator is provided by shunting the circuit with a suitably chosen value of resistance in series with a thermistor. Networks of resistors and thermistors are used as compressors, expanders and limiters in transmission circuits. [F H B.]

Bibliography: J. A. Becker, C. B Green, G. L. Pearson, Properties and uses of thermistors—thermally sensitive resistors, *Trans. AIEE*, 65:711–725, 1946, Victory Engineering Corp., *Thermistors Data Book*

Thermoanalysis

A group of techniques for continuously measuring the effects attending chemical or physical changes caused by various processes that occur in a single or multicomponent system as the temperature is varied at a selected heating rate A linearly increasing temperature cycle of 5–15°C per minute is commonly used, and the measurements are conducted in an environment of air or other static or dynamically controlled atmosphere at elevated, ambient, or reduced pressures The resulting thermograms describe the system uniquely in terms of the heat effects, changes in weight, volume of gas evolved or absorbed, or one of many other characteristic changes in physicochemical properties and reactions that occur as functions of temperature The principal thermoanalytical methods are thermogravimetry (TG) and differential thermal analysis (DTA). Instrumentation required for these complementary techniques includes the sample holder contained in a furnace equipped with a temperature programmer, and a means for converting the physicochemical property being measured into an electrical signal that can be measured as a function of the temperature, or of time, con-

s, which are analogous to cooling curves, can ...rned from DTA curves plotted as a function ...creasing temperature. For a discussion of the ...ication of adiabatic temperature difference ...urements to volumetric analysis, *see* TITRA-

See also EQUILIBRIUM, PHASE; THERMAL EX...SION; TRANSITION POINT. [S.C.]

...*bliography:* C. Duval, *Inorganic Thermogravi...ric Analysis*, 1953; W. J. Smothers, *Differential ...rmal Analysis*, 1957.

...ermochemistry

...ranch of physical chemistry dealing with the ...t effects which accompany chemical reactions, ...formation of solutions, and changes in the ...sical state of substances, such as the fusion of a ...stalline solid or the vaporization of a liquid. ...hen these processes evolve heat, they are said to ...exothermic. Conversely, those absorbing heat ...e endothermic. A knowledge of the magnitudes ...such heat effects is of practical value to the en...neer in solving problems of heating and refriger...ion and in controlling chemical reactions at suit...ble temperatures. Such data are also of great the...retical importance to the scientist who wishes to ...lculate the chemical affinity or free energy for ...arious reactions, hypothetical or real. *See* FREE ...NERGY, THERMODYNAMICS (CHEMICAL)

These heat effects are usually measured in calo...ries, partly for historical reasons and partly be...cause the figures thereby obtained are of more con...venient magnitude than when the joule, the fundamental unit of energy, is employed The pres...ent-day calorie is arbitrarily defined as being equal to 4.184 absolute joules, rather than in terms of the heat capacity of water, which varies with tempera...ture.

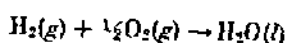
Fundamental concepts. The principle of conser-

...crease in its internal energy ΔE is related to the heat q absorbed by the system from its surround...ings and the work w done by the system on these surroundings by the equation

$$\Delta E = q - w$$

In practice, a negative value for any of these quan...tities is common and simply denotes a decrease in the energy content of the system, or evolution of heat, or work done upon the system, as the case may be. If the change under consideration takes place at a constant temperature and also at con...stant volume so that w has zero value $\Delta E = q$ and may be termed the heat of reaction at constant volume

The chemist and engineer, however, are often in...terested in processes which take place at constant pressure, frequently 1 atm. In this case, w will be equal to the change in volume ΔV in such a proc...ess multiplied by this pressure P provided no other work is done by, or on, the system Then the pre...ceding equation may be rearranged to yield



$\Delta E + P\,\Delta V = q$. It is now convenient to introduce a new function, enthalpy, which is defined as $H = E + PV$, and accordingly here $\Delta H = q$. Thus, the change in enthalpy measures the heat effect in a process at constant temperature and pressure. Because these are the conditions most frequently encountered in practice, ΔH is a more generally useful value than ΔE in a reaction. Of course, since they differ simply by the $P\,\Delta V$ term, either can be computed from the other

Experimental results. Many experimental determinations of these heat effects (that is, ΔE and ΔH values) in chemical reactions and other transformations have been made by various calorimetric methods during the past 100 years Most of the results of the measurements by the pioneers, such as J Thomsen and M P. E. Berthelot, involved considerable uncertainties because generally their chemicals were impure, and their apparatus often was crude Since 1930, however, there has been a notable renascence in thermochemical studies. It has been characterized by the use of extremely pure chemical substances and great improvements in calorimetric equipment and procedures. Consequently, present-day measurements of the heats of combustion of organic compounds, for instance, usually involve uncertainties under 0.1% and in some cases, under 0.02%. The studies of F. D. Rossini and collaborators at the National Bureau of Standards are outstanding examples of such highly accurate work *See* CALORIMETRY

In certain cases, it is also possible to evaluate ΔH for a process by indirect, noncalorimetric methods Thus, some reactions, especially in the field of inorganic chemistry, can be effected in connection with reversible galvanic cells From a careful measurement of the electromotive force in such cases at several different temperatures, the ΔH value can be calculated by the Gibbs-Helmholtz equation For a number of reactions, extremely accurate data have been obtained by this method. Likewise a measurement of the equilibrium constants for a reaction at two or more different temperatures can be used to deduce a moderately accurate value, sometimes at fairly elevated temperatures, by calculation with the van't Hoff equation. Similarly many heats of vaporization have been derived indirectly from vapor pressure measurements and the Clausius-Clapeyron equation. *See* VAPOR PRESSURE

These heat effects are frequently recorded in a shorthand fashion by writing a chemical equation for the process involved and the corresponding ΔE or ΔH value, depending on whether the process occurs at constant volume or constant pressure. Thus, the equation for the reaction between gaseous hydrogen and oxygen to produce a gram-mole of liquid water at 25°C (298°K) is

$$H_2(g) + \tfrac{1}{2}O_2(g) \rightarrow H_2O(l)$$
$$\Delta H_{298} = -68{,}317 \pm 10\ \text{cal}$$

Here the abbreviations in parentheses indicate that

establish quantitatively the high-temperature reactions and thermal stabilities of solid or liquid substances, identify intermediate reaction products, reveal the sequence of reactions that accompany the various reactions (decomposition or degradation), and ultimately reveal the kinetics of these reactions, from which reaction mechanisms may be deduced. A typical thermogravimetric curve is shown in Fig. 2*a*.

Differential thermal analysis. This is an experimental method derived from thermal analysis which involves measuring the temperature difference between a substance and a thermally inert reference compound (commonly aluminum oxide) as they are simultaneously heated to elevated temperatures at a predetermined rate. Series-connected thermocouples or thermistors (Fig. 3) are used to detect the temperature differences that occur as the sample undergoes enthalpic changes caused by chemical and physical reactions; these are manifested by the absorption or evolution of heat. Because the sample temperature lags behind that of the inert reference material during endothermal processes, and exceeds it when exothermal reactions occur, the DTA curves, which are plotted as functions of increasing temperature, consist of bands and peaks corresponding to the characteristic physicochemical reactions of the substance. Endotherms are plotted in a downward direction as illustrated by the DTA curve for calcium nitrate tetrahydrate (Fig 2c)

Among the endothermal physical reactions that may be detected are crystalline transitions (both first- and second-order), fusion, vaporization or sublimation, and desorption. Those of a chemical nature include dehydration, desolvation, therm decomposition, some metatheses, and many sol state reactions. Exothermal phenomena inclu certain crystalline transitions, solidification, a sorption and absorption, hydration, corrosion, d rect reaction of solids or liquids with gaseo oxidants, certain metatheses, and some solid-sta reactions.

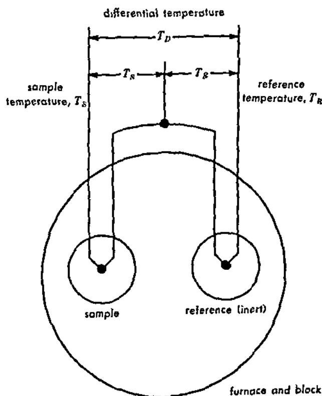


Fig 3. Schematic diagram illustrating the principles of the apparatus for differential thermal analysis.

Analytical applications of this technique inclu the identification and characterization of ma types of materials: metals, alloys, minerals, i organic and organic compounds, and synthetic well as naturally occurring polymeric system This is based upon the uniqueness of their respe tive DTA curves, that is, the composite of te peratures and rates at which various reactions o cur, the number and sequential appearance of bo discrete and overlapping bands and peaks, as we as their relative magnitudes and areas. These are are determined by many parameters, the princip ones being the relative thermal effects involve that is, the amount of heat associated with th processes detected, the size, granulation, and for of the sample, heating rate, use of either a stat or dynamically controlled atmosphere, geomet of the apparatus, and the sensitivity of the mea uring and recording system The amount of a give material in a mixture can be estimated quantit tively from the height of a peak or band, the rel tive area under a given band, or the slope of th DTA band which is formed at a relatively constar temperature during crystalline transition or fusion This procedure requires the calibration of the sys tem for the given experimental conditions.

Differential enthalpic analysis. This method i used to evaluate the changes in heat content whic are continuously detected in the sample by DTA and internally compensated for by electrical heat ing of either the substance or the inert referenc compound, each of which is mixed with an elec trically conductive powder such as carbon. A quan titative measure of the particular enthalpic proc ess is the electrical power required to maintain th system in a state of thermal equilibrium during th chemical or physical reaction.

Related thermoanalytical techniques involve measuring the volume or pressure of gas evolved or absorbed, under conditions of constant pressure or volume, respectively, changes in electrical conductivity, length, or specific volume of the substance; or the continuous variation in any other chemical or physical property as a function of temperature. The value of these types of measurements is greatly enhanced when derivative thermoanalytical techniques are employed. Electronic or mechanical means, as well as the simultaneous analysis of two identical samples which are maintained at a slightly different temperature, may be used to obtain curves which are the calculus-type derivative of the primary curve, that is, the rate of change of the property such as weight (DTG) (Fig. 2*b*) or temperature difference (DDTA), recorded as a function of temperature or of time. Phase dia-

rmed out at different temperatures, often much gher The scientist and engineer, therefore, may sh to know the ΔH° value for this second temrature T_2 in a case where he has the correspondg value for T_1, which is frequently 298°K. For is problem, Kirchhoff's law may be utilized. This w dates from 1858 and can be considered as a orollary of the principle of conservation of engy It may be stated in equation form

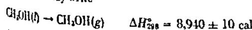
$$\Delta H_{T_2} = \Delta H_{T_1} + \int_{T_1}^{T_2} \Delta C_p \, dT$$

here ΔC_p represents the difference between the eat capacities at constant pressure of the products and initial substances in the reaction The use f this equation obviously requires adequate heat apacity data over the temperature range involved nd a knowledge of the enthalpy change at one emperature. In practice, the ΔH° values for many processes vary considerably with temperature. For example, in the case of the formation of methane

ticularly when only liquids or solids are concerned However, it can be of considerable importance in certain gas-phase reactions, such as the ammonia synthesis, where pressures of hundreds of atmospheres may be used and the substances then exhibit considerable deviations from ideal gas behavior Suitable thermodynamic equations can be utilized for estimating such changes in enthalpy with pressure.

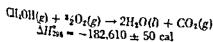
Changes in physical state. So far, the discussion of heat effects has been concerned entirely with chemical reactions. However, changes in physical state, such as the vaporization of a liquid or the

und to the gas as a result of a calorimetric determination, one may write



$CH_3OH(l) \rightarrow CH_3OH(g) \qquad \Delta H^\circ_{298} = 8{,}940 \pm 10$ cal

Then, in accordance with Hess's law, this thermochemical equation can be combined with that previously given for the combustion of liquid methanol to yield a value for the combustion of the gaseous form.



$$CH_3OH(g) + \tfrac{3}{2}O_2(g) \rightarrow 2H_2O(l) + CO_2(g)$$
$$\Delta H^\circ_{298} = -182{,}610 \pm 50 \text{ cal}$$

Such a result is valuable for developing the thermodynamic treatment of the methanol synthesis

Solutions. The thermochemistry of solutions can be touched upon only briefly here. The heat effects

tion.

By contrast, the solution processes for many inorganic substances in water are frequently exothermic as a result of the hydration of ions and other departures from ideality. Two heat quantities commonly occur in such cases, the integral and the differential heats of solution. The integral heat of solution is the ΔH per mole of solute when it is dissolved in a given solvent, such as water, to form a solution of a particular concentration On the other hand, the differential heat of solution is the ΔH effect when 1 mole of the solute is dissolved in such a large volume of solution of this particular concentration that the concentration is not appreciably changed These two ΔH values become identical for an infinitely dilute solution. However, they differ by 2000 cal in the case of a 5.0 molal solution of aqueous sulfuric acid. See SOLUTION.

[G.S.P.]

Bibliography: F. D. Rossini, Modern thermochemistry, *Chem. Rev.*, 18:233, 1936, F. D Rossini et al., *Selected Values of Chemical Thermodynamic Properties*, U.S Nat. Bur. Standards Circ. 500, 1952; L. E. Steiner, *Introduction to Chemical Thermodynamics*, 2d ed., 1948.

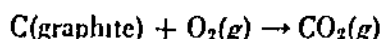
Thermocline

A layer of sea water in which the temperature decrease is greater than that of the overlying and underlying water Such layers are semipermanent features of the oceanic temperature structure, and their depth and thickness show marked variation with season, latitude and longitude, and local environmental conditions Since the three-dimensional temperature structure has a great effect on many oceanic properties, such as the transmission of sound, the study of the nature and behavior of the thermocline is of extreme importance to many oceanographic interests, both economic and military In general, two major types of thermocline may be identified, the permanent thermocline and the seasonal thermocline In addition to these types, shallow thermoclines or similar stable layers often occur, owing to diurnal heating of the surface waters See UNDERWATER SOUND.

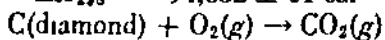
Permanent thermocline. This feature is so named because its character is virtually unchanged seasonally In Arctic and Antarctic regions, the water is cold from top to bottom As this dense water flows south and north, respectively, it sinks beneath warmer water which moves outward from the Equator This gives rise to the temperature discontinuity known as the permanent thermocline. The cold water flowing slowly through the deep ocean basins exhibits conservative properties throughout all the oceans, however, on top of this dense layer lie a number of shallow layers whose character varies from ocean to ocean The top of the permanent thermocline is quite shallow at the

the hydrogen and oxygen are gases, and the water produced is liquid. The subscript after ΔH represents the absolute temperature for the reaction and the degree symbol indicates that each substance involved is under the standard pressure of 1 atm. The negative value for this ΔH indicates that heat is evolved. The uncertainty in this result is only 10 cal, because this determination was made with extreme accuracy at the U.S. National Bureau of Standards. Because it applies to the process of forming a compound from its elements, this quantity is the standard heat of formation.

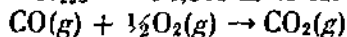
Similarly, a few additional thermochemical equations for typical combustion reactions can be shown:



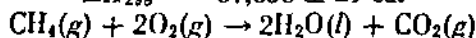
$$C(\text{graphite}) + O_2(g) \rightarrow CO_2(g) \qquad \Delta H^\circ_{298} = -94{,}052 \pm 11 \text{ cal}$$



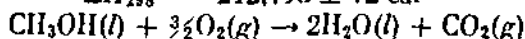
$$C(\text{diamond}) + O_2(g) \rightarrow CO_2(g) \qquad \Delta H^\circ_{298} = -94{,}505 \pm 23 \text{ cal}$$



$$CO(g) + \tfrac{1}{2}O_2(g) \rightarrow CO_2(g) \qquad \Delta H^\circ_{298} = -67{,}636 \pm 29 \text{ cal}$$



$$CH_4(g) + 2O_2(g) \rightarrow 2H_2O(l) + CO_2(g) \qquad \Delta H^\circ_{298} = -212{,}798 \pm 72 \text{ cal}$$

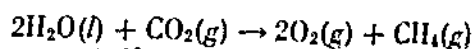


$$CH_3OH(l) + \tfrac{3}{2}O_2(g) \rightarrow 2H_2O(l) + CO_2(g) \qquad \Delta H^\circ_{298} = -173{,}670 \pm 50 \text{ cal}$$

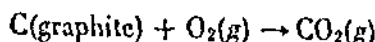
The first two equations involve the production of carbon dioxide from graphite and diamond, respectively It is interesting to note that these two crystalline forms of carbon have appreciable differences in their combustion values Because the end product for each combustion is the same, this means that the enthalpy content of diamond is 453 cal higher per gram-atom than that of graphite. For practical purposes, graphite is usually taken as the standard form of this element. The ΔH values for the combustion of carbon monoxide, methane, and methanol are data of great industrial importance which were quite uncertain for many years

Law of Hess. It is frequently desirable to calculate the heat effect in a particular reaction for which a direct experimental determination is not available. The thermochemist then has recourse to the law of Hess This is essentially a corollary of the principle of conservation of energy, although it was discovered a few years earlier, in 1840, by direct experimentation. According to this law, the net heat effect or change in enthalpy in taking a chemical system from a state A to a state B must be the same regardless of whether the path between these two states is traversed directly in one step or in a roundabout fashion by two or more steps.

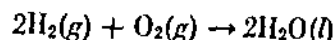
Thus, the law of Hess permits the combination of chemical equations and the corresponding ΔH values to arrive at the ΔH value for the desired reaction. This may now be illustrated by a computation of the heat of formation of methane from its elements. The preceding combustion values can be used. Here, however, the equation for the combustion of methane must be written backward, also reversing the ΔH sign



$$2H_2O(l) + CO_2(g) \rightarrow 2O_2(g) + CH_4(g) \qquad \Delta H^\circ_{298} = 212{,}798 \pm 72 \text{ cal}$$

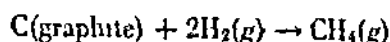


$$C(\text{graphite}) + O_2(g) \rightarrow CO_2(g) \qquad \Delta H^\circ_{298} = -94{,}052 \pm 11 \text{ cal}$$



$$2H_2(g) + O_2(g) \rightarrow 2H_2O(l) \qquad \Delta H^\circ_{298} = -136{,}634 \pm 20 \text{ cal}$$

Algebraic addition of these thermochemical equations yields



$$C(\text{graphite}) + 2H_2(g) \rightarrow CH_4(g) \qquad \Delta H^\circ_{298} = -17{,}888 \pm 75 \text{ cal}$$

The uncertainty in this final ΔH value has been computed by probability methods from the uncertainties in the contributing data.

Similar calculations may be made for many compounds where it is impractical, or even impossible, to make a direct calorimetric evaluation of the heats of formation. All such calculations deal with the changes in enthalpy accompanying changes in chemical composition or physical state. The absolute values of the enthalpies of the elements or of the resulting compound are not known; such knowledge is not really essential for the practical purposes of the scientist or engineer. It is merely the enthalpy difference that is important. Accordingly, the enthalpy of each element is arbitrarily assigned a zero value in a standard reference state at all temperatures This reference state is usually the most stable form of the element at room temperature and atmospheric pressure. Thus, carbon is taken as β-graphite, sulfur as the rhombic crystalline form, and the three elements, hydrogen, oxygen, and nitrogen, in the form of diatomic gas at 1 atm pressure. With this convention, extensive tables of ΔH° values for the formation of various substances, usually at 25°C and 1 atm, have been developed Noteworthy among such compilations are the tables of *Selected Values* issued by the National Bureau of Standards and the American Petroleum Institute.

From these tables of data, the ΔH°_{298} values may now be derived for numerous reactions, some of which are purely hypothetical. One needs only to subtract the enthalpy values for the initial materials from the corresponding value of the reaction product Thus, for an important reaction for the production of ethyl alcohol from ethylene and water, one finds



$$\underset{(12{,}496)}{C_2H_4(g)} + \underset{(-68{,}317)}{H_2O(l)} \rightarrow \underset{(-66{,}356)}{C_2H_5OH(l)} \qquad \Delta H^\circ_{298} = -10{,}535 \text{ cal}$$

by subtracting the enthalpies of formation of ethylene and water from that for the alcohol These contributing values are here placed directly below the respective compounds in the chemical equation

Effects of temperature and pressure. Up to this point, only the heat effects in processes at 25°C have been considered. This temperature has become a standard for the recording of thermal data. However, in practice many chemical reactions are

rked metals are necessary for good serv- (7) cost.

nf temperature curves for six widely-used uples are illustrated in Fig. 1 and the emperature limits and corrosion character- e listed in the table. Application condi- etermine the thermocouple used. When are applicable, it is common to select the the greatest sensitivity and the least cost. emf generated by a thermocouple does not upon the size of the wire. Therefore, small, e thermocouples are ideally suited for the ement of temperature in small spaces or rapid temperature changes require good ic characteristics. Time constants of the tude of 1 second are possible in gases at pheric pressure moving with a moderate ty. On many applications, however, an ade- life can be achieved only with a heavier wire xample, 14 gage) and one or two protecting , both of which cause a reduction in speed of nse. At low temperatures (1000°F and be- , a single metallic well is generally adequate gh temperatures, the thermocouple is often ted in a primary tube of porcelain or fused a, and this is protected by a secondary tube of al, silicon carbide, or fire clay. The construc- s used are always a compromise between serv- life, dynamic characteristics and errors due to

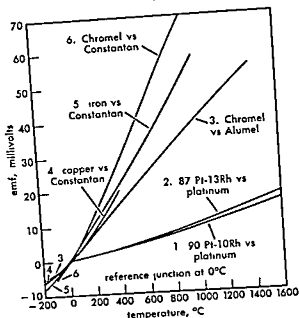


Fig 1. Temperature-thermal emf curves for common types of thermocouples

ice to maintain the reference junction at a temperature slightly above atmospheric by a miniature thermostatically controlled oven and to zero the instrument by adding the corresponding voltage. Other instruments allow the reference-junction temperature to vary with ambient conditions within the voltage measuring instrument and provide a calibrated electrical or mechanical manual adjustment so that with proper adjustment the instrument reading is correct. In recent years most instruments that are used with only one type of

emperature limits and corrosion characteristics of thermocouples*

Positive element	Negative element	Temperature range, °C	Influence of temperature and gas atmospheres
90% Pt 10% Rh, 87% Pt 13% Rh	Platinum	0 to 1450	Resistance to oxidizing atmosphere very good, resistance to reducing atmosphere poor; platinum corrodes easily above 1000°C, should be used in gastight ce-
Chromel-P	Alumel	−200 to 1100	
Iron	Constantan	−200 to 750	Oxidizing and reducing atmospheres have little effect on accuracy, best used in dry atmospheres, resistance to oxidation good up to 400°C, poor above 700°C, resistance to reducing atmosphere good up to 400°C; protection from oxygen, moisture, and sulfur required
Copper	Constantan	−200 to 350	Subject to oxidation and alteration above 400°C due to copper, above 600°C due to Constantan wire, contamination of copper affects calibration greatly; resistance to oxidizing atmosphere good; resistance to reducing atmosphere good; requires protection from acid fumes
Chromel-P	Constantan	−100 to 1000	Chromel attacked by sulfurous atmosphere, resistance to oxidation good; resistance to reducing atmosphere poor

* From D. M. Considine (ed.), *Process Instruments and Controls Handbook*, McGraw-Hill, 1957.

Equator, reaches maximum depth at mid-latitudes, and becomes shallow again at about 50° latitude. The thermocline disappears between 55 to 60°N or S. In general, as the permanent thermocline deepens, it becomes thicker and the temperature gradient within it decreases. Figure 1 indicates schematically variations with latitude in the characteristics of the permanent thermocline. *See* SEA WATER.

Seasonal thermocline. This feature is a summer phenomenon found at shallower water depth than the permanent thermocline in all the world's oceans except those perennially ice-infested. As air temperatures rise above ocean temperatures in the spring season and the sea surface receives more heat than it loses by radiation and convection, the surface water begins to warm so that a negative temperature gradient develops in the first few feet (Fig. 2*a*). The surface waters are then mixed by

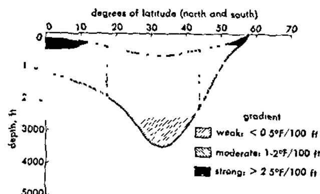


Fig. 1. Character of the permanent thermocline. Based on average figures for depth, thickness, and gradient within thermocline

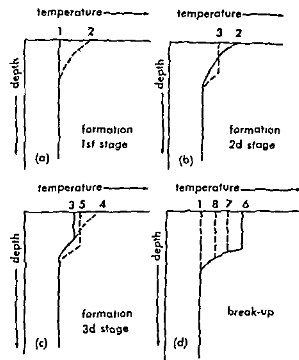


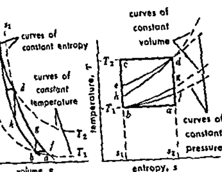
Fig. 2. (a–d) Formation and break-up of seasonal thermocline. Numbers show sequence in development and disappearance of thermocline; profiles show temperature structure.

transfer of energy from the wind. Although this mixing serves to lower the surface temperature, the net effect is a downward transport of heat and formation of an isothermal layer whose temperature is warmer than the underlying water (Fig. 2*b*). A strong temperature gradient, or seasonal thermocline, is thus formed between the isothermal surface layer and water beneath. This process repeats itself (Fig. 2*c*) until the gradient in the seasonal thermocline becomes so strong that summer winds cannot impart sufficient energy to drive the isothermal layer deeper. From July through September such a surface layer of mixed water underlain by a strong negative temperature gradient is found in most of the ocean. As air temperatures fall in autumn, the water loses heat to the atmosphere by convective and radiative processes, and the surface layer is cooled to the temperature of the water below. The seasonal thermocline breaks up (Fig. 2*d*), to form again the following spring. Seasonal thermoclines may be affected locally by vertical wind mixing, currents, and heat exchange across the interface between ocean and atmosphere. Further distortions may occur because the density discontinuity associated with thermoclines provides a favorable environment for internal waves. Practically all physical processes occurring in the sea have an effect on thermocline characteristics. *See* WAVE (INTERNAL); *see also* HALOCLINE. [J.J.SC.]

Thermocouple

A device that uses the voltage developed by the junction of two dissimilar metals to measure temperature difference. Two dissimilar wires welded together at one end form the basic thermocouple. The junction is used as the sensing portion and is placed at the point where temperature is to be measured. The other ends of the wires are maintained at a known reference temperature. Voltage developed across the junction is roughly proportional to the temperature difference between the junction and the opposite ends of the wires. If the free ends are connected, a direct current will flow through the circuit. The direction of the current depends on whether the junction temperature is higher or lower than the reference temperature. For a thorough discussion of the basic principles of the thermocouple, *see* THERMOELECTRICITY.

While all dissimilar metals exhibit the thermoelectric effect, only a few are in wide use. The major characteristics which make certain metals or combinations of metals outstanding for this purpose are (1) stability or reproducibility, the emf does not change rapidly with time; (2) constant or controllable composition, small impurities or changes in composition of wire from end to end or lot to lot can result in varying or nonidentical emf curves; (3) corrosion resistance, wires do not deteriorate or change properties in oxidizing or reducing atmospheres; (4) sensitivity, the emf generated per degree temperature change is large; (5) range, the couple can be used through a broad range of temperatures; (6) ruggedness, tough but



cycles are, in the order of decreasing efficiency
Carnot cycle (a-b-c-d-a)
Brayton cycle (b-e-d-f-b)
Diesel cycle (b-e-d-g-b)
Otto cycle (b-h-d-g b)

parison of principal thermodynamic cycles

ve engine illustrates this arrangement of the ic processes, called an open cycle because new s enters the system boundaries and the spent ust leaves it.

he basic processes of the cycle, either open or ed, are heat addition, heat rejection, expansion, compression These basic processes are always sent in a cycle even though there may be dif-ences in working substance, the individual proc-

rious power cycles by using an ideal system such the air-standard cycle as illustrated This is an eal, frictionless mechanism enveloping the sys-m, with a permanent unit charge of air behaving

re described as either a heat transfer or a me-hanical work term At the end of a series of proc-sses, the state of the system is the same as it was nitially. Because no chemical changes take place within the system, the same unit air charge is con-ceivably capable of going through the cyclic proc-esses repeatedly

Whereas this air-standard cycle is an idealiza-tion of an actual cycle, it provides an amenable method for the introductory evaluation of any power cycle Its analysis defines the upper limits of performance toward which the actual cycle per-formance may approach It defines trends, if not absolute values for both ideal and actual cycles The air-standard cycle can be used to examine such cycles as the Carnot and those applicable to the automobile engine the diesel engine, the gas tur-bine, and the jet engine [J B.]

Bibliography G A Hawkins, *Thermodynamics*, 2d ed, 1951, J H Keenan, *Thermodynamics*, 1941, J F Lee and F W Sears, *Thermodynamics*, 1955

Thermodynamic principles

Laws governing the conversion of energy to and from heat and the methods employed for such transformations are the subject matter of thermo-dynamics.

In this article, underlying principles are re-viewed. For detailed treatment and application of these principles, see THERMODYNAMIC CYCLE; THERMODYNAMIC PROCESSES. For discussion of spe-cific aspects of thermodynamics, see CARNOT CY-CLE; HEAT; STEAM; THERMOCHEMISTRY; THERMO-ELECTRICITY.

Forms of energy. The capacity of matter for producing an effect is its energy. The effect may be in the form of mechanical work, biological growth, or other manifestations. For convenience,

ther in its state of aggregation or in the motion of the matter, it is stored. Energy which is associ-ated with matter by virtue of the position of the matter is classed as potential energy.

Potential energy arises from the position of a body relative to an external datum level, such as a skull cracker raised to the top of a wrecker's crane; and from the internal state of a body, such as a spring stretched beyond its static length. Po-tential energy is measured by the work it can per-form.

The energy stored in a given quantity of matter by virtue of its motion is termed kinetic energy. A bullet approaching a target at high velocity pos-sesses kinetic energy, as evidenced by its ability to penetrate wood blocks and steel plates. An automo-bile moving along a level highway has kinetic en-ergy of translation, while a moving wheel on a ma-chine tool has kinetic energy of rotation. There are other types of energy associated with electric and magnetic fields, chemical energy, and energy con-tained within the atomic nucleus.

Internal energy refers to the energy stored in a molecular system. The molecular world consists of enormous numbers of extremely minute entities called molecules and atoms. Everywhere in this world there is varied activity. It might be thought that no laws govern the traffic in the molecular world; everything appears to happen as a matter of chance. Although it is not yet feasible to study the motion of a single molecule, statistical meth-ods make it possible to predict the behavior of large aggregates of molecules.

The stored energy in a molecular system consists of kinetic energy of translation, rotation, and vi-bration, and of potential energy due to the bond-ing forces in the system In general, the engineer is not concerned with the separation of the total energy of a molecular system into the component parts; hence the total stored energy is called the internal energy.

The internal energy of gases, liquids, and sol-ids can be computed. As an example, the internal

thermocouple have automatic reference-junction compensation, and their scale or chart is calibrated directly in temperature units. Figure 2 shows some typical circuit connections for thermocouples.

Extension leads are commonly used to connect the measuring junction with the reference junction when the connections are greater than 20 ft in length. Extension wire is alloyed to match the thermoelectric effect of the thermocouple wires through the limited temperature ranges to which the lead wires are subjected. The extension wire has better mechanical characteristics, lower resistance, or lower cost than the thermocouple wire. When long lead wires to the measuring instrument are necessary, it is sometimes desirable to locate the reference junction near the measuring junction and make the meter connection by ordinary copper wire.

Any method of measuring small voltages accurately may be used for thermocouple voltages. The millivoltmeter and the potentiometer in various forms are in wide use. Precision measurements with thermocouples (say to 0.1°F) are possible in certain ranges with specially calibrated units. On industrial applications, errors of 5°F and larger occur, but the actual magnitude depends upon the thermocouple materials, the temperature level, the lead wires, the compensation, the voltage-measuring system, and the installation. Although thermocouple systems have many limitations, they are particularly advantageous for applications requiring remote indication or recording, for those on which the measuring junction must be replaced relatively frequently, and for those requiring measurements in the temperature range between 800 and 2400°F. In the laboratory and in experimental work, the thermocouple is frequently a convenient substitute for the more accurate, but less rugged, liquid-in-glass thermometer.

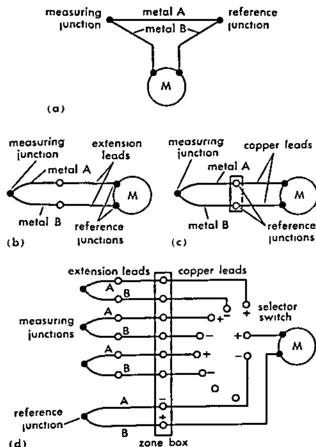


Fig. 2. Typical thermocouple circuits (a) Single thermocouple circuit requiring no extension leads. (b) Single thermocouple circuit having the reference junctions at a distance from the measuring instrument (c) Single thermocouple circuit having the reference junctions at the measuring instrument. (d) Multiple thermocouple circuit having the reference junctions at a distance from the measuring instrument.

The thermopile is a number of thermocouples connected in series or parallel The series circuit provides a higher sensitivity (greater emf per degree) than one thermocouple alone and is often selected for this reason. Either the series or parallel circuit may be used for obtaining an indication which approaches the average of the several temperatures

For other means of measuring temperature, *see* TEMPERATURE MEASUREMENT. [R.E.CL.]

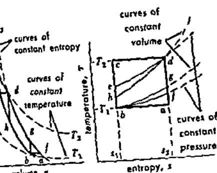
Bibliography: D. M. Considine (ed), *Process Instruments and Controls Handbook*, 1957.

Thermodynamic cycle

In a thermodynamic cycle, one form of energy, such as heat from the combustion of a fuel, is in part converted to another form such as mechanical energy on a shaft, and the remainder is rejected as low-grade heat.

Common features of cycles. A thermodynamic cycle requires, in addition to the supply of incoming energy, (1) a working substance, usually a gas or vapor, (2) a mechanism in which the processes can be carried through sequentially, and (3) a thermodynamic sink to which the residual heat can be rejected The cycle itself is a repetitive series of operations

There is a basic pattern of processes common to power-producing cycles. There is a compression process wherein the working substance undergoes an increase in pressure and therefore density. There is an addition of thermal energy (*see* HEAT) from a source such as a fossil fuel, a fissile fuel, or solar radiation There is an expansion process during which work is done by the system on the surroundings. There is a rejection process where thermal energy is transferred to the surroundings. The algebraic sum of the energy additions and abstractions is such that some thermal energy is converted into mechanical work. A steam cycle embracing a boiler, a prime mover, a condenser, and a feed pump is typical of this cyclic arrangement in which the thermodynamic fluid, steam, is used over and over again. An alternative, after the net work flows from the system, is to employ a change of mass within the system boundaries, the spent working substance being replaced by a fresh charge ready to repeat the cyclic events. The auto-



ycles are, in the order of decreasing efficiency
Carnot cycle (a-b-c-d-a)
Brayton cycle (b-e-d-j-b)
Diesel cycle (b-e-d-g-b)
Otto cycle (b-h-d-g-b)

arison of principal thermodynamic cycles.

e engine illustrates this arrangement of the c processes, called an open cycle because new enters the system boundaries and the spent ust leaves it.

he basic processes of the cycle, either open or ed, are heat addition, heat rejection, expansion, compression. These basic processes are always sent in a cycle even though there may be differences in working substance, the individual processes, pressure ranges, temperature ranges, mechanisms, and heat transfer arrangements

Air-standard cycle. It is convenient to study the rious power cycles by using an ideal system such the air standard cycle as illustrated This is an leal, frictionless mechanism enveloping the system, with a permanent unit charge of air behaving accordance with the perfect gas relationships

The unit air charge is assumed to have an initial state at the start of the cycle to be analyzed. Each process is assumed to be perfectly reversible, and all effects between the system and the surroundings are described as either a heat transfer or a mechanical work term At the end of a series of processes, the state of the system is the same as it was initially Because no chemical changes take place within the system, the same unit air charge is contenably capable of going through the cyclic processes repeatedly

Whereas this air-standard cycle is an idealization of an actual cycle, it provides an amenable method for the introductory evaluation of any power cycle Its analysis defines the upper limits of performance toward which the actual cycle performance may approach It defines trends, if not absolute values for both ideal and actual cycles. The air standard cycle can be used to examine such cycles as the Carnot and those applicable to the automobile engine the diesel engine, the gas turbine, and the jet engine. [J A.]

Bibliography G A Hawkins, *Thermodynamics*, 2d ed., 1951 J H Keenan, *Thermodynamics*, 1941 J F Lee and F W Sears, *Thermodynamics*, 1955

Thermodynamic principles

Laws governing the conversion of energy to and from heat and the methods employed for such transformations are the subject matter of thermodynamics.

In this article, underlying principles are reviewed. For detailed treatment and application of these principles, *see* THERMODYNAMIC CYCLE; THERMODYNAMIC PROCESSES. For discussion of specific aspects of thermodynamics, *see* CARNOT CYCLE; HEAT; STEAM; THERMOCHEMISTRY; THERMOELECTRICITY.

Forms of energy. The capacity of matter for producing an effect is its energy. The effect may be in the form of mechanical work, biological growth, or other manifestations. For convenience, energy is subdivided into two general classes: stored energy and energy in transition.

Stored energy. When energy resides in matter either in its state of aggregation or in the motion of the matter, it is stored. Energy which is associated with matter by virtue of the position of the matter is classed as potential energy.

Potential energy arises from the position of a body relative to an external datum level, such as a skull cracker raised to the top of a wrecker's crane; and from the internal state of a body, such as a spring stretched beyond its static length. Potential energy is measured by the work it can perform.

The energy stored in a given quantity of matter by virtue of its motion is termed kinetic energy. A bullet approaching a target at high velocity possesses kinetic energy, as evidenced by its ability to penetrate wood blocks and steel plates. An automobile moving along a level highway has kinetic energy of translation, while a moving wheel on a machine tool has kinetic energy of rotation. There are other types of energy associated with electric and magnetic fields, chemical energy, and energy contained within the atomic nucleus.

Internal energy refers to the energy stored in a molecular system The molecular world consists of enormous numbers of extremely minute entities called molecules and atoms. Everywhere in this world there is varied activity. It might be thought that no laws govern the traffic in the molecular world; everything appears to happen as a matter of chance. Although it is not yet feasible to study the motion of a single molecule, statistical methods make it possible to predict the behavior of large aggregates of molecules.

The stored energy in a molecular system consists of kinetic energy of translation, rotation, and vibration, and of potential energy due to the bonding forces in the system. In general, the engineer is not concerned with the separation of the total energy of a molecular system into the component parts; hence the total stored energy is called the internal energy.

The internal energy of gases, liquids, and solids can be computed. As an example, the internal

energy of a perfect gas is independent of the gas volume and is dependent only on the absolute temperature of the gas. This statement is referred to as Joule's law.

Energy in transition. Clearly distinguished from the various forms of stored energy is energy in transition. Work is defined as a form of energy in transit resulting from a force acting through a given distance. To distinguish clearly between work and other forms of stored energy, consider a tank of compressed gas. By releasing the gas through a nozzle, its stored potential energy is converted into kinetic energy of the moving molecules. The conversion is brought about in this case by the pressure in the tank. The conversion could also be brought about by adding heat at the nozzle. This would cause the molecules to emerge from the nozzle at a greater velocity than that caused by pressure alone. The kinetic energy in the molecules can then be further transformed by directing them against a turbine blade. The molecules give up their kinetic energy to the turbine blade to rotate the turbine and do useful work.

In many systems, force is not constant. For example, when the gas expands in the cylinder its pressure and volume both vary; hence to determine the work term it is necessary to evaluate the integral of the product of pressure and the differential change in volume

If two bodies at different temperatures are placed together, energy transfers from the hotter to the colder body. The energy in transit between bodies at different temperatures is termed heat

The modes by which heat may be transferred are conduction, convection, and radiation Conduction is the passage of heat through adjacent layers of matter as a result of the transfer of energy in the molecular structure because of a temperature difference The flow of heat through the wall of a furnace is an example of this type of heat transfer Heat is transferred by convection in a fluid by the movement of quantities of material from a high temperature region to one at a lower temperature The actual motion of the fluids may result from differences in density, which is known as natural or free convection, or the transfer may be produced by mechanical devices such as pumps and fans, in which case it is called forced convection Heat may be transferred from a hot body to a colder body even though no material substance connects them This type of transfer is accomplished by electromagnetic radiation and is termed radiant heat transfer.

A transformation such as the vaporization of a liquid, the condensation of a vapor, the melting of a solid, the freezing of a liquid, or the transition of a solid from one crystalline form to another is accompanied by a transfer of heat. The quantity of heat transferred to or from a unit quantity of matter during the transformation is the latent heat.

Measures of energy. Energy can be transformed from one form to another, therefore heat energy can be expressed in terms of its equivalent mechanical energy. In dealing with thermodynamic systems, other measures of energy are more directly applicable. But regardless of its form or the way it is measured, energy can neither be created nor destroyed; the total energy associated with a thermodynamic conversion remains constant. This is the First Law of Thermodynamics. (If relativistic transformations are introduced, the law is extended to include the equivalence of mass and energy.) Basic to all such measures of energy is power

Power, referring to a mechanical system, is defined as the rate of performing work, or the quotient of the work performed divided by the time elapsed during the performance Power in electrical transformations is the rate of doing work against electric charges in electric fields Typical units for mechanical and electrical power are horsepower and watt.

Enthalpy is a property used in the analysis of a thermodynamic system. The enthalpy per unit mass of material, or the specific enthalpy, is numerically equal to the sum of the internal energy per unit quantity and the product of the pressure and the specific volume (see ENTHALPY)

Thermal capacity is defined as the heat transferred to a unit quantity of matter which is sufficient to cause a 1° change in the temperature of the body. The ratio of the heat transferred to the temperature change of the material is the mean thermal capacity Specific heat of a material at a given temperature is the ratio of the thermal capacity of the substance at that temperature to the thermal capacity of water at a standard temperature and pressure. So defined, specific heat is a dimensionless quantity. If the standard temperature and pressure are selected so that the thermal capacity of water is unity, then the specific heat of a substance is numerically equal to its thermal capacity. Because of this equality the terms "specific heat" and "thermal capacity" are often used interchangeably Specific heat of a substance is dependent on the temperature and also on the process involved In the case of gases, two important specific heats are those for constant-volume and constant-pressure processes A definite relationship exists between these two specific heats which is of importance in the analyses of systems involving gases. Values of specific heat for some simple gases can be derived from kinetic theory, which takes into account the degrees of freedom and the law of equipartition of energy That is, internal energy of a dynamical system in thermal equilibrium is equally divided between the various degrees of freedom. For more complex gases, specific heat data are obtained from spectroscopic data and analyses.

Law of degradation of energy. An analysis of the Carnot cycle indicates the maximum amount of transferred heat that can be converted into useful work, the remainder being unavailable as work. To convert heat into work, energy is transferred from a high-temperature source to the heat engine. The

gine converts some of that energy into work and rejects the remainder to a lower temperature reservoir. No deviations from this procedure have been observed, and thus it appears that another natural law exists: the law of degradation of energy.

The eternal struggle to smooth out variations in energy levels is one of the most pervasive forces on the earth—the force activating the changes in composition, placement, temperature, and physical movements of matter. Man cannot prevent the leveling out of energy levels or potentials, but he can, by taking pains, direct slight transformations, and by so doing he can make the flow of energy do some of his chores. Applied to systems in which heat is transferred to an apparatus and part of it is converted into work, the law of degradation of energy is usually referred to as the Second Law of Thermodynamics. Many statements of the second law have been formulated. One, by Max Planck, follows: "It is impossible to construct an engine which, working in a complete cycle, will produce no effect other than the raising of a weight and the cooling of a heat reservoir." The amount of energy which is unavailable in a heat engine for performing work, and which must be dumped on the energy scrap pile is an important item in the analysis of thermodynamic systems.

An important state property called entropy is used extensively to evaluate this phenomenon. Entropy is a measure of the unavailable energy of a process. The concept of entropy can also be developed from probability studies (*see* ENTROPY).

Equations of state. The physical forms of matter are divided into three phases: solid, liquid, and gas. The First and Second Laws of Thermodynamics are general in character and may be used in the analysis of a system; however, additional information is needed. This may be acquired from the equations of state, which are relations among variables used to define the state of a material under equilibrium conditions. The present extent of knowledge does not permit the formulation of a general equation of state for all materials. The engineer and the scientist use equations based on kinetic theory, and information obtained by experiment. Over wide ranges of conditions, several equations of state may be required. The state of a gas can change in several ways.

Gas laws. Robert Boyle published his remarkable researches on the properties of air at ordinary temperature in 1662. As a result of this work, a gas law bearing his name evolved. According to Boyle's law, for a given quantity of gas at a fixed temperature, the product of pressure and volume is constant. In other words, at fixed temperature, pressure varies inversely with volume. Later experiments show that this law is not valid over all ranges of pressure and temperature. At low pressures, the law predicts the behavior of most gases satisfactorily.

Jacques A. Charles, John Dalton, and Joseph L. Gay Lussac later conducted investigations dealing with the change in volume of a gas during heating under constant pressure conditions, which resulted in additional information. Charles' law states that, with the pressure of a gas remaining constant, *change in volume* is accompanied by a proportional change in temperature. Similarly, with the volume of a gas remaining constant, a change in pressure is accompanied by a corresponding change in absolute temperature.

A characteristic gas equation may be developed on the basis of Boyle's law and Charles' law, according to which the pressure of any given quantity of gas is proportional to the absolute temperature and inversely proportional to the volume. Or, the product of the pressure and volume is equal to the product of a constant and the absolute gas temperature. This relation is based on the assumption that the gases follow the laws of Boyle and Charles. Because, under certain conditions, gases do not follow these laws, it is expected that the characteristic gas equation does not accurately represent the behavior of gases for all ranges of pressure and temperature. Even though the relation is only approximate, it is extremely useful for the solution of many problems. For real vapors and gases at relatively low pressures and high temperatures, the relation represents the behavior of gases and vapors with good accuracy. The relation may also be developed from kinetic theory. Ideal gas or perfect gas are terms used to describe a hypothetical gas which would follow the characteristic equation under all conditions. The equation of state for an ideal or perfect gas is the characteristic relation developed from Boyle's and Charles' laws. The constant in the equation is called the ideal gas constant and is fixed for a given gas but varies from one gas to another. The individual gas constants may be replaced by the ratio of a universal gas constant and the molecular weight. In this case, the universal gas constant is the same for all gases. The ideal gas law states that the product of the pressure and volume is equal to the product of the *universal gas constant and absolute temperature* divided by the molecular weight.

When the deviation of an actual gas from the ideal gas law becomes excessive, other relations must be used to avoid errors in the analysis of systems. The two most common procedures are to employ more complicated equations of state or to modify the ideal gas law by means of a correction factor known as the compressibility factor. This factor is dependent on pressure, temperature, and type of gas. Charts from which compressibility factors may be obtained are available.

In the development of the ideal gas law from kinetic theory, *two assumptions are made:* (1) the various molecules of the gas are considered to have negligible volume; (2) the force of attraction between adjacent molecules is considered extremely small. In 1873, J. D. van der Waals presented his equation of state, which took into account the finite size of the molecule and the attractive forces acting between molecules. Because actual molecules

occupy space, van der Waals reasoned that the volume occupied by a real gas would be less than that of the ideal gas. The pressure of the gas must be adjusted for the attractive forces between molecules. The equation is called van der Waal's equation of state for gases. Although the equation does not represent values of the properties of gases with a high degree of accuracy, it is of considerable use in studying the properties of gases. It is a refinement of the ideal gas equation. *See* VAN DER WAALS EQUATION.

Many equations of state have been proposed since van der Waals presented his. One of these is the Beattie-Bridgeman equation of state. This is a complex refinement of the earlier equations. *See* BEATTIE-BRIDGEMAN EQUATION.

Investigators have measured the volume of a fixed mass of gas at constant temperature over a wide range of pressures. Results are usually expressed in the form of the product of the pressure and specific volume equal to a series equation of increasing powers of pressure. Coefficients for the various terms are called virial coefficients. Considerable research has been conducted to refine the coefficients, so that the properties may be determined with greater accuracy.

Mixtures of gases. Properties for mixtures of gases are obtained from the characteristic equations for an ideal gas and Avogadro's law, Dalton's law, and Amagat's law. Avogadro's law states that equal volumes of ideal gases at the same temperature and pressure contain the same number of molecules According to Dalton's law or the law of additive pressures, the total pressure exerted by a mixture of ideal gases in equilibrium equals the sum of the partial pressures. The law of additive volumes, formulated by E H Amagat and A. Leduc and commonly known as Amagat's law, states that for a mixture of ideal gases, the total volume is equal to the sum of the volumes which the constituent gases would occupy at the total pressure of the mixture, and at the same temperature This law has been found to hold for a mixture of real gases with a fair degree of accuracy. [G.A H.]

Bibliography· G. A Hawkins, *Thermodynamics*, 2d ed., 1951; J. H. Keenan, *Thermodynamics*, 1941; H. C. Weber, and H P Meissner, *Thermodynamics for Chemical Engineers*, 1957, M W. Zemansky, *Heat and Thermodynamics*, 1957.

Thermodynamic processes

When any property of an aggregation of matter and energy changes, a process takes place. When there are thermal effects, the change is a thermodynamic process. The participants in a process are first identified as a system to be studied, the boundaries of the system are established, the initial state of the system is determined, the path of the changing states is laid out; and finally, supplementary data are stated to establish the thermodynamic process. These steps will be taken in the following paragraphs.

A system and its boundaries. To evaluate the results of a process, it is necessary to know the participants that undergo the process and t mass and energy. A region, or a system, is sele for study, and its contents determined. This re may have both mass and energy entering or l ing during a particular change of conditions, these mass and energy transfers may result changes both within the system and within the roundings which envelop the system.

As the system undergoes a particular chang conditions, such as a balloon collapsing becaus the escape of gas, or a liquid solution brought t boil in a nuclear reactor, the transfers which oc in the system's mass and energy can be evaluatec the boundaries of the arbitrarily defined syst under analysis

A question that immediately presents itself whether a system such as a tank of compressed should have boundaries which include or excl the metal walls of the tank. The answer deper upon the aim of the analysis. If its aim is to est lish a relationship among the physical propert of the gas, such as to determine how the pressu of the gas varies with the gas temperature at c stant volume, then only the behavior of the gas involved; the metal walls do not belong within t system However, if the problem is to determi how much externally applied heat would be quired to raise the temperature of the enclosed g [illegible] through which the heat flows to reach the gaseo contents In the laboratory, regardless of whe the system boundaries are taken, the walls will a ways play a role, and must be reckoned with.

State of a system. To establish the exact pa of a process, the initial state of the system must b determined. This initial thermodynamic state c condition of the system is characterized by a def nite pressure, temperature, volume, and other suc dimensions. If a dimension is reproducible, it called a property of the system If enough prop erties of a system are determined, the state of th system is determined; the remaining physical prop erties may be established based upon this state o the system The number of properties required t specify the state of a system depends upon th complexity of the system Whenever a system changes from one state to another then a process occurs.

After an initial state of a system is established, several questions may be asked. What makes a system change from this state to another? What driving potentials cause a process to occur, or cause a change of state of the system? Also, how far forward will a process go? These questions will be approached by first discussing the fundamental aspects of the equilibrium of a system.

Whenever an unbalance occurs in an intensive property such as temperature, pressure, or density, either within the system or between the system and its surroundings, the force of the unbalance initiates a process that causes a change of state. Examples of an unbalanced potential property that can

tiate a change of state are the unequal molecular concentration of different gases within a single rigid enclosure, a difference of temperature across the system boundary; a difference of pressure normal to a nonrigid system boundary; or a difference of electrical potential across the system boundaries. The direction of the change of state caused by the unbalanced force is such as to reduce the unbalanced driving potential. All changes of state tend to decelerate as this driving potential is decreased.

Equilibrium. The decelerating rate of change implies that all states move toward new static conditions of equilibrium. When there are no longer any unbalanced forces acting within the boundaries of a system, or between the system and its surroundings, then no mechanical changes can take place, and the system is said to be in mechanical equilibrium. A system in mechanical equilibrium, such as a mixture of hydrogen and oxygen, might under certain conditions undergo a chemical change. If, however, a chemical change is not possible under any condition, as in a mixture of neon and argon, then the mixture is said to be in chemical as well as mechanical equilibrium.

If all parts of a system in chemical and mechanical equilibrium attain a uniform temperature, and [illegible]

Whenever a system is in mechanical, chemical and thermal equilibrium, so that no mechanical, chemical, or thermal changes can occur, the system is in thermodynamic equilibrium. The particular state of thermodynamic equilibrium reached by a [illegible] s or di [illegible] , rather [illegible] condition.

Process path. If under the influence of an unbalanced intensive factor, the state of a system is altered, then the change of state of the system is described in terms of the end states, or of the difference between the initial and final physical properties.

The path of a change of state is the locus of the whole series of states through which the system passes when going from an initial to a final state. It is possible to duplicate the path that a system follows by any of several different processes. For example, [illegible] along this definite path may be produced either by heat transfer alone, by vigorously stirring the liquid, or by any combination of the two. Thus, the exact process causing a change of state must be particularized by more than merely presenting the path traversed, and must be described by the method which induces the change. To be adequate in defining the process, the description of the method must include at least the heat or the work entering the system during the process.

From the above descriptions of systems, boundaries, states, and processes, there are several corollaries. First, all properties are identical for identical states. Second, the change in a property between initial and final states is independent of path or processes. The third corollary is that a quantity whose change is fixed by the end states and is independent of the path is a point function or a property.

Pressure-volume-temperature diagram. Whereas the state of a system is a point function, the change of state of a system, or a process, is a path function. Various processes or methods of change of a system from one state to another may be depicted graphically as a path on a plot using thermodynamic properties as coordinates.

It is determined by Gibbs' phase rule, and verified by experience, that a pure substance, which is either in the liquid or in the gaseous phase, in the absence of motion, gravity, capillarity, electricity and magnetism, has only two independent thermodynamic properties. Among the properties of a thermodynamic substance which can be quantitatively evaluated are the pressure, temperature, specific volume, internal energy, enthalpy, entropy, viscosity, and electrical resistivity. From among these properties, any two may be selected. If these two prove to be independent of each other, when the values of these two properties are fixed, the state is determined, and the values of all the other properties are also fixed.

There are several variable properties which are frequently and conveniently measured: pressure, volume, and temperature, any two being the selected independent variables, and the third being the dependent variable. To depict the relationship among these physical properties of the particular working substance, these three variables may be used as the coordinates of a three-dimensional space. The resulting surface is a graphic presentation of the equation of state for this working substance, and all possible equilibrium states of the substance lie on this p-v-T surface. The p-v-T surface may be extensive enough to include all three

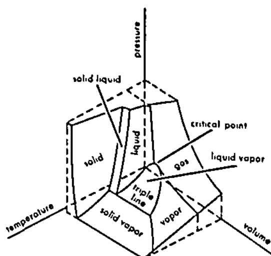


Fig. 1 Portion of p-v-T surface for a typical substance.

occupy space, van der Waals reasoned that the volume occupied by a real gas would be less than that of the ideal gas. The pressure of the gas must be adjusted for the attractive forces between molecules. The equation is called van der Waal's equation of state for gases. Although the equation does not represent values of the properties of gases with a high degree of accuracy, it is of considerable use in studying the properties of gases. It is a refinement of the ideal gas equation. *See* VAN DER WAALS EQUATION.

Many equations of state have been proposed since van der Waals presented his. One of these is the Beattie-Bridgeman equation of state. This is a complex refinement of the earlier equations. *See* BEATTIE-BRIDGEMAN EQUATION

Investigators have measured the volume of a fixed mass of gas at constant temperature over a wide range of pressures. Results are usually expressed in the form of the product of the pressure and specific volume equal to a series equation of increasing powers of pressure. Coefficients for the various terms are called virial coefficients. Considerable research has been conducted to refine the coefficients, so that the properties may be determined with greater accuracy.

Mixtures of gases. Properties for mixtures of gases are obtained from the characteristic equations for an ideal gas and Avogadro's law, Dalton's law, and Amagat's law. Avogadro's law states that equal volumes of ideal gases at the same temperature and pressure contain the same number of molecules. According to Dalton's law or the law of additive pressures, the total pressure exerted by a mixture of ideal gases in equilibrium equals the sum of the partial pressures. The law of additive volumes, formulated by E H Amagat and A. Leduc and commonly known as Amagat's law, states that for a mixture of ideal gases, the total volume is equal to the sum of the volumes which the constituent gases would occupy at the total pressure of the mixture, and at the same temperature This law has been found to hold for a mixture of real gases with a fair degree of accuracy. [C A H]

Bibliography: G A. Hawkins, *Thermodynamics*, 2d ed., 1951; J. H. Keenan, *Thermodynamics*, 1941; H C. Weber, and H. P. Meissner, *Thermodynamics for Chemical Engineers*, 1957; M. W. Zemansky, *Heat and Thermodynamics*, 1957.

Thermodynamic processes

When any property of an aggregation of matter and energy changes, a process takes place When there are thermal effects, the change is a thermodynamic process. The participants in a process are first identified as a system to be studied; the boundaries of the system are established; the initial state of the system is determined; the path of the changing states is laid out; and finally, supplementary data are stated to establish the thermodynamic process. These steps will be taken in the following paragraphs.

A system and its boundaries. To evaluate the results of a process, it is necessary to know the participants that undergo the process and their mass and energy. A region, or a system, is selected for study, and its contents determined. This region may have both mass and energy entering or leaving during a particular change of conditions, and these mass and energy transfers may result in changes both within the system and within the surroundings which envelop the system.

As the system undergoes a particular change of conditions, such as a balloon collapsing because of the escape of gas, or a liquid solution brought to a boil in a nuclear reactor, the transfers which occur in the system's mass and energy can be evaluated at the boundaries of the arbitrarily defined system under analysis.

A question that immediately presents itself is whether a system such as a tank of compressed air should have boundaries which include or exclude the metal walls of the tank The answer depends upon the aim of the analysis. If its aim is to establish a relationship among the physical properties of the gas, such as to determine how the pressure of the gas varies with the gas temperature at constant volume, then only the behavior of the gas is involved; the metal walls do not belong within the system. However, if the problem is to determine how much externally applied heat would be required to raise the temperature of the enclosed gas

[illegible]

through which the heat flows to reach the gaseous contents. In the laboratory, regardless of where the system boundaries are taken, the walls will always play a role, and must be reckoned with.

State of a system. To establish the exact path of a process, the initial state of the system must be determined. This initial thermodynamic state or condition of the system is characterized by a definite pressure, temperature, volume, and other such dimensions If a dimension is reproducible, it is called a property of the system. If enough properties of a system are determined, the state of the system is determined, the remaining physical properties may be established based upon this state of the system The number of properties required to specify the state of a system depends upon the complexity of the system. Whenever a system changes from one state to another then a process occurs.

After an initial state of a system is established, several questions may be asked. What makes a system change from this state to another? What driving potentials cause a process to occur, or cause a change of state of the system? Also, how far forward will a process go? These questions will be approached by first discussing the fundamental aspects of the equilibrium of a system.

Whenever an unbalance occurs in an intensive property such as temperature, pressure, or density, either within the system or between the system and its surroundings, the force of the unbalance initiates a process that causes a change of state. Examples of an unbalanced potential property

ples of an intensive property are pressure, temperature, and voltage; extensive ones are volume, current, and mass. Thus, in differential form, work has been presented as the product of a pressure exerted against an area which sweeps through an infinitesimal volume

$$dW = P\,dV$$

Similarly, the energy stored in a wire stretched an infinitesimal amount is

$$dW = -F\,dl$$

where F is tensile force and l is elongation.

By extending this approach, one can depict transferred heat as the product of an intensive property, temperature, and a distributed or extensive property defined as entropy, for which the symbol is s.

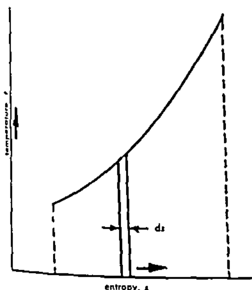


Fig. 5. Heat transferred during a process is area under path in T-s plane

If an infinitesimal quantity of heat Q is transferred during a reversible process, this process may be expressed mathematically as

$$dQ = T\,ds$$

with dQ standing for the infinitesimal quantity of transferred heat, T the absolute temperature, and ds the infinitesimal entropy quantity.

Further, a plot of the change of state of the system undergoing this reversible heat transfer can be drawn on a plane in which the coordinates are absolute temperature and entropy (Fig. 5). The total heat transferred during this process equals the area between this plotted line and the horizontal axis.

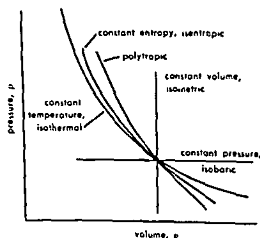
Reversible processes. Not all energy contained in or associated with a mass can be converted into useful work. Under ideal conditions, only a fraction of the total energy present can be converted into work. The ideal conversions which retain the maximum available useful energy are reversible processes.

Characteristics of a reversible process are that the working substance is always in thermodynamic equilibrium, and the process involves no dissipative effects such as viscosity, friction, inelasticity, electrical resistance, or magnetic hysteresis. Thus, reversible processes proceed quasistatically so that the system passes through a series of states of thermodynamic equilibrium, both internally and with its surroundings. This series of states may be traversed just as well in one direction as in the other.

If there are no dissipative effects, all work done by the system during a process in one direction can be returned to the system during the reverse process. When such a process is reversed so that the system returns to its starting state, it leaves no evidence of the events on the surroundings or in the working substance after it returns to its initial state.

It is impossible to satisfy these conditions of a quasistatic process with no dissipative effects; a reversible process is an ideal abstraction which, although useful for theoretical calculations, nevertheless, is not realizable in practice.

There are five significant reversible processes (see ISENTROPIC PROCESS; ISOBARIC PROCESS; ISO-



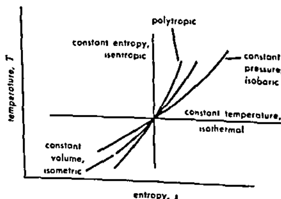


Fig. 6. P-v and T-s plots of reversible processes for a perfect gas.

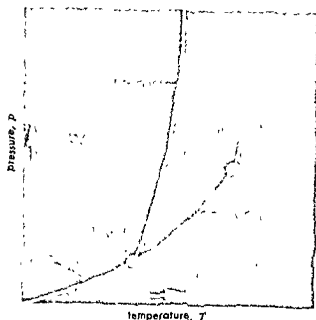


Fig. 2. Portion of equilibrium surface projected on p-T plane.

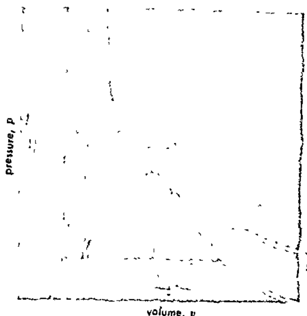


Fig 3. Portion of equilibrium surface projected on p-v plane.

phases of the working substance: solid, liquid, and vapor

Figure 1 is a schematic of a portion of the p-v-T surface for most real substances; it is characterized by contraction of the substance on freezing. Water is one of the few exceptions to this condition; it expands upon freezing, and its resulting p-v-T surface is somewhat modified where the solid and liquid phases abut.

Because a p-v-T surface represents all equilibrium conditions of the working substance, any line on the surface represents a possible reversible *process, or a succession of equilibrium states.*

Work of a process. One can *project the three-*dimensional surface onto the p-T plane or onto the p-v plane to get Figs. 2 and 3 respectively The plot on the p-v plane has a special significance: the area under any reversible path on this plane represents the work done during the process The fact that this p-v area represents useful work can be demonstrated by the following example

Let a fluid undergo an infinitesimal expansion in a cylinder equipped with a frictionless piston, and let this expansion perform useful work on the surroundings The work done during this infinitesimal expansion is the force multiplied by the distance through which it acts

$$dW = F\,dl$$

wherein dW is an infinitesimally small work quantity, F the force, and dl the infinitesimal distance through which F acts.

But force F is equal to the pressure P of the fluid times the area A of the piston, or PA. However, the product of the area of the piston times the infinitesimal displacement is really the infinitesimal volume swept by the piston, or $A\,dl = dV$, with dV equal to an infinitesimal volume. Therefore

$$dW = PA\,dl = P\,dV$$

The work term is found by integration as

$$_1W_2 = \int_1^2 P\,dV$$

Figure 4 shows that the integral represents the area under the path described by the expansion from state 1 to state 2 on the p-v plane. Thus, the *area on the p-v plane represents work done during this expansion process*

Temperature-entropy diagram. Energy quantities may be depicted as the product of two factors: an intensive property, and an extensive one. Exam-

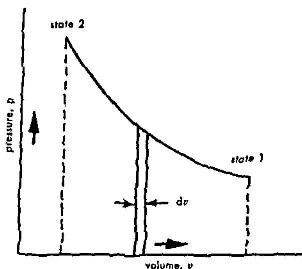


Fig. 4. Area under path in p-v plane is work.

…quasi-static is often used instead of reversible. …atural, spontaneous processes are irreversible; …rsible performance represents for real proc…s a limiting condition, which in some cases, …example, the operation of certain electrochemi…cells, may be approached quite closely. Because …ntaneous processes lead toward equilibrium, any …nitesimal process taking place at equilibrium …st be reversible. The problem of establishing …eria for spontaneous change and equilibrium …n becomes one of criteria for irreversibility …d reversibility.

First law of thermodynamics. Let an adiabatic …ield be defined as one which will restrict the …teraction of the system and the surroundings to …e performance of work on one by the other. (A [illegible]

…e work required to accomplish a given change in [illegible]

…s an extensive property of the state of the system. …this function is called the internal energy, E, and …thus, for any adiabatic process,

$$\Delta E = E_2 - E_1 = -W_{ad}$$

Here E_2 and E_1 represent the values of the internal energy for the initial and final states, and W_{ad} the work done by the system. *See* INTERNAL ENERGY, WORK

Suppose now the same change in state is effected under nonadiabatic conditions while the previously implied restriction to a closed system is maintained. The work W will no longer be equal to ΔE, and there may be defined a quantity Q such that

$$\Delta E = Q - W$$

Here, Q is termed the heat transferred between the system and the surroundings. This intuitively appealing relation cannot be used satisfactorily by itself as a statement of the first law because heat can properly be defined only in terms of the internal energy change and work accompanying a process. For a differential process (infinitesimal change in state) there results

$$dE = \delta q - \delta w$$

In this equation, dE represents the exact differential of the internal energy function, the symbols δq and δw are used to emphasize that these are inexact differentials in general. Heat and work have no meaning relative to a single state, and may be termed path or process functions; the internal energy, in contrast, has a definite value for any single state of the system.

It is necessary to adopt sign conventions to designate the direction of flow of heat or work. The relations written above correspond to using a positive sign for heat absorbed by, or work done by, the system, a negative sign for heat transferred to or work done by the surroundings.

Consider now a closed, static (no bulk motion) system under ordinary conditions such that pressure is the only mechanical force acting. The only work done is that corresponding to a change in the volume of the system, and $\delta w = p\,dv$. For constant volume, as indicated by the subscript used below, the first law then gives

$$dE = \delta q$$

or for a finite change in state

$$\Delta E = Q_v$$

The heat effect Q_v accompanying this constant volume process is then directly equal to the internal energy change for the system. This relation has important thermochemical applications in constant volume oxygen bomb calorimetry. *See* CALORIMETRY

For constant pressure, under the above specified restrictions there results

$$dE = \delta q_p - p\,dv \quad \text{or} \quad dE + p\,dv = \delta q_p$$

If there is defined a new variable, the enthalpy, $H = E + PV$, then

$$dH = dE + p\,dv + v\,dp$$

At constant pressure, $dH = dE + p\,dv$, so that $dH = \delta q_p$. For a finite process at constant pressure, the heat effect Q_p is equal to the enthalpy change of the system, ΔH. *See* ENTHALPY.

Since E, P, and V are all state functions, the enthalpy must be also, and the change ΔH in enthalpy for a process again is independent of the path and is fixed by the initial and final states. This principle is exploited in thermochemistry, where it forms the basis for the indirect evaluation of heats of reaction. It should be noted, however, that the relation $\Delta H = Q_p$ is not general, but is restricted to cases where only PV work is done. It does not hold, for example, for the operation of an electrochemical cell, where electrical work is done also. *See* THERMOCHEMISTRY.

The term heat capacity refers to the proportionality factor between the heat δq absorbed by a system and the change dt in empirical temperature which results:

$$\delta q = C\,dt$$

Heat capacities may be defined in terms of any temperature scale desired, but in thermodynamics, there is employed the Kelvin absolute temperature scale, designated by T, whose existence is established through the second law of thermodynamics. *See* HEAT CAPACITY; TEMPERATURE.

For a single homogeneous phase of fixed mass and composition, and with only pressure-volume work possible, the thermodynamic heat capacity at

METRIC PROCESS; ISOTHERMAL PROCESS; POLYTROPIC PROCESS). In four of these, one parameter is constant (Fig. 6). All are idealized processes and are for a closed or nonflow system consisting of an ideal gas.

Irreversible processes. Actual processes of a gas deviate from the idealized situation of a quasi-static process devoid of dissipative effects. The extent of the deviation from ideality is correspondingly the extent of the irreversibility of the process.

Real expansions take place in finite time, not infinitely slowly; and these expansions occur with friction of rubbing parts, turbulence of the fluid, pressure waves sweeping across and rebounding through the cylinder, and finite temperature gradients driving the transferred heat. These dissipative effects, the kind of effects that make pendulums and yo-yos slow down and stop, also make the work output of actual irreversible expansions less than the ideal work of a corresponding reversible process.

Vapor process. Figure 1 presents a three-dimensional surface in pressure-volume-temperature coordinates. This surface contains all *equilibrium* states that the working substance can occupy. In the portion of the surface outside the *liquid-vapor* dome, where temperature is high and pressure is low, the working substance behaves as though it were a perfect gas. Attractive forces appear to be negligible between any two of the molecules in a low-pressure, rarified gas.

In the region just outside the dome, the gas is referred to as a vapor. When a gaseous system follows a process path which approaches the liquid-vapor dome on the thermodynamic surface (that is, when higher pressures or lower temperatures are attained), attractive forces appear among the more crowded, slower-moving molecules, and the behavior differs from that of a rarified gas. If the path of the process cuts through the liquid-vapor dome, there is a change of phase, and some of the vapor condenses to form liquid droplets in the gaseous vapor.

If heat is added at constant pressure to a mixture of liquid droplets and steam vapor, the resulting constant pressure process includes an expansion of the vapor, doing work, as the droplets of liquid vaporize at a constant temperature and pressure. When no liquid droplets are left, additional heat input expands the vapor, but this expansion is now accompanied by a temperature rise, producing a superheated vapor.

If a rigid tank is filled with such a superheated vapor, and if heat is removed from the tank, the system undergoes a constant volume process. No work is performed in such a process, but the transferred heat is exactly equal to the reduction in internal energy stored in the system. This constant-volume cooling causes the vapor to lose both pressure and temperature until the state path crosses the saturation dome. After that, further cooling produces not only a reduction of pressure and te[mperature], but also some condensation and "rai[n]" within the rigid tank. [J.I...]

Bibliography: G. A. Hawkins, *Thermodynami[cs]*, 2d ed., 1951; J. H. Keenan, *Thermodynami[cs]*, 1941; J. H. Keenan and F. G. Keyes, *Thermod[y]namic Properties of Steam*, 1936; J. F. Lee a[nd] F. W. Sears, *Thermodynamics*, 1955; H. C. Web[er] and H. P. Meissner, *Thermodynamics for Chemic[al] Engineers*, 2d ed., 1957; M. W. Zemansky, *He[at] and Thermodynamics*, 3d ed., 1951.

Thermodynamics (chemical)

The application of thermodynamic principles [to] problems of chemical interest, notably the estab[lishment] of criteria for spontaneous change an[d] equilibrium in chemical systems, the effects [of] variation of experimental conditions on the proper[ties] of materials, and the energy changes ac[companying] physical and chemical processes. Th[e] relations employed are developed from three gen[eral] principles, the first, second, and third laws o[f] thermodynamics, which are postulates summarizin[g] conclusions drawn from experimental evidenc[e] obtained in the study of material systems. Thes[e] laws are not considered subject to proof, thei[r] acceptance is based on the validity of the result[s] thereby obtained.

Basic concepts. Thermodynamics is commonly concerned with the interaction of a system (the material of interest) with its surroundings (everything else). An isolated system undergoes no interaction with its surroundings; for an open system, the interaction may include an exchange of matter, whereas for a closed system it cannot. Any system will have a number of interrelated properties, such as pressure, volume, and number of moles of various constituents, the term state of the system refers to a condition characterized by a particular set of mutually compatible values for these properties. The number of independent variables which must be specified to identify a particular state of a system depends on the complexity of the system and the experimental conditions to which it is subjected, but will never be less than two, even for fixed mass and composition.

Two kinds of processes are recognized in thermodynamics, reversible and irreversible. After a reversible process has occurred, both system and surroundings can be restored simultaneously to their initial states, following an irreversible process this is possible for one of the two, but not both. If a process is considered as resulting from an unbalance in forces between the system and surroundings (for example, an unbalance in pressure producing an expansion of a gas), for reversible performance the unbalance must be infinitesimal in magnitude, for the irreversible case finite During the execution of a reversible process, the system is thus only infinitesimally removed from being in a state of equilibrium with its surroundings. The

term quasi-static is often used instead of reversible.

Natural, spontaneous processes are irreversible; reversible performance represents for real processes a limiting condition, which in some cases, for example, the operation of certain electrochemical cells, may be approached quite closely. Because spontaneous processes lead toward equilibrium, any infinitesimal process taking place at equilibrium must be reversible. The problem of establishing criteria for spontaneous change and equilibrium then becomes one of criteria for irreversibility and reversibility.

First law of thermodynamics. Let an adiabatic shield be defined as one which will restrict the interaction of the system and the surroundings to the performance of work on one by the other. (A Dewar flask constitutes one good approximation to such a shield.) The first law of thermodynamics states that for a system enclosed in such a shield, the work required to accomplish a given change in state depends only on the initial and final states of the system, and is independent of the path by which

[illegible]

this function is called the internal energy, E, and thus, for any adiabatic process,

$$\Delta E = E_2 - E_1 = -W_{ad}$$

Here E_2 and E_1 represent the values of the internal energy for the initial and final states, and W_{ad} the work done by the system. *See* INTERNAL ENERGY, WORK.

Suppose now the same change in state is effected under nonadiabatic conditions, while the previously implied restriction to a closed system is maintained. The work W will no longer be equal to ΔE, and there may be defined a quantity Q such that

$$\Delta E = Q - W$$

e, Q is termed the heat transferred between the em and the surroundings. This intuitively ap-ling relation cannot be used satisfactorily by lf as a statement of the first law because heat properly be defined only in terms of the in-nal energy change and work accompanying a cess. For a differential process (infinitesimal ange in state), there results

$$dE = \delta q - \delta w$$

this equation, dE represents the exact differen-l of the internal energy function, the symbols and δw are used to emphasize that these are exact differentials in general. Heat and work have o meaning relative to a single state, and may be rmed path or process functions; the internal nergy, in contrast, has a definite value for any ingle state of the system.

It is necessary to adopt sign conventions to designate the direction of flow of heat or work. The relations written above correspond to using a positive sign for heat absorbed by, or work done by, the system, a negative sign for heat transferred to or work done by the surroundings.

Consider now a closed, static (no bulk motion) system under ordinary conditions such that pressure is the only mechanical force acting. The only work done is that corresponding to a change in the volume of the system, and $\delta w = p\,dv$. For constant volume, as indicated by the subscript used below, the first law then gives

$$dE = \delta q$$

or for a finite change in state

$$\Delta E = Q_v$$

The heat effect Q_v accompanying this constant volume process is then directly equal to the internal energy change for the system. This relation has important thermochemical applications in constant-volume oxygen bomb calorimetry. *See* CALORIMETRY.

For constant pressure, under the above specified restrictions there results

$$dE = \delta q_p - p\,dv \quad \text{or} \quad dE + p\,dv = \delta q_p$$

If there is defined a new variable, the enthalpy, $H = E + PV$, then

$$dH = dE + p\,dv + v\,dp$$

At constant pressure, $dH = dE + p\,dv$, so that $dH = \delta q_p$. For a finite process at constant pressure, the heat effect Q_p is equal to the enthalpy change of the system, ΔH. *See* ENTHALPY.

Since E, P, and V are all state functions, the enthalpy must be also, and the change ΔH in enthalpy for a process again is independent of the path and is fixed by the initial and final states. This principle is exploited in thermochemistry, where it forms the basis for the indirect evaluation of heats of reaction. It should be noted, however, that the relation $\Delta H = Q_p$ is not general, but is restricted to cases where only PV work is done. It does not hold, for example, for the operation of an electrochemical cell, where electrical work is done also. *See* THERMOCHEMISTRY.

The term heat capacity refers to the proportionality factor between the heat δq absorbed by a system and the change dt in empirical temperature which results:

$$\delta q = C\,dt$$

Heat capacities may be defined in terms of any temperature scale desired, but in thermodynamics, there is employed the Kelvin absolute temperature scale, designated by T, whose existence is established through the second law of thermodynamics. *See* HEAT CAPACITY; TEMPERATURE.

For a single homogeneous phase of fixed mass and composition, and with only pressure-volume work possible, the thermodynamic heat capacity at

constant volume, C_v, is defined through the equation

$$\delta q_v = C_v\, dT$$

Since $\delta q_v = dE$,

$$C_v = \left(\frac{\partial E}{\partial T}\right)_v$$

Similarly, the heat capacity at constant pressure C_p is given by $C_p = (\partial H/\partial T)_p$. These heat capacities are extensive properties of the state of the system.

Second law of thermodynamics. It it found impossible to formulate criteria for spontaneous change and equilibrium in terms of the functions E or H only. These problems, and that of the maximum efficiency with which heat can be converted into work, can be solved through the second law of thermodynamics, which may be stated, following Max Planck, as follows: "It is impossible to construct a machine which will operate in a complete cycle and produce no effects other than to do work and exchange heat with a single reservoir."

The operation of a typical heat engine may be discussed in terms of the Carnot cycle in which the engine absorbs heat Q_2 isothermally from a reservoir at empirical temperature t_2, and rejects heat Q_1 at constant empirical temperature t_1; the processes leading between these two levels are adiabatic, and hence involve no heat exchange with the surroundings. For a complete cycle, there can be no change in internal energy for the engine system, and thus the work done per cycle must be equal to $Q_2 - Q_1$. The efficiency of conversion of heat into work is then given by

$$\epsilon = \frac{W}{Q_2} = (Q_2 - Q_1)/Q_2 = 1 - (Q_1/Q_2)$$

Irreversibility in operation can only decrease the efficiency below the maximum possible value, which then corresponds to reversible performance. All reversible engines operating between two specified temperature levels must have the same efficiency in both directions, since otherwise there could be constructed a composite machine, consisting of a more efficient engine driving a less efficient one in reverse, which could violate the second law by virtue of having the heat delivered to the former pumped from the lower temperature reservoir by the latter, at the expenditure of less work than that delivered by the more efficient engine This combination could thus deliver work to the surroundings in cyclic operation, while exchanging heat with a single reservoir only. The efficiency of a reversible engine can thus depend only on the two temperature levels between which it works. It may further be shown that

$$\frac{Q_2}{Q_1} = \frac{f(t_2)}{f(t_1)}$$

where $f(t_2)$ depends only on the level t_2, and $f(t_1)$ only on the level t_1. The ratio Q_2/Q_1 thus provides a relative ordering of the two temperature levels in the absolute sense, since the only assumption is of reversible operation with no specification of any working fluid. The Kelvin absolute temperature scale is obtained simply by setting $f(t_2)$ equal to the temperature T_2 on this scale, and $f(t_1) = T_1$. The thermodynamic efficiency ϵ of the engine then becomes $\epsilon = 1 - (T_1/T_2)$. The experimental evaluation of temperatures on this scale is actually accomplished by gas thermometry. *See* CARNOT CYCLE; HEAT PUMP; TEMPERATURE MEASUREMENT.

Entropy. The second law also requires the existence of a system of a new state function called the entropy. Let heat in amount δq be transferred to a closed system at Kelvin temperature T. Then, if the value of $\delta q/T$ is added for each differential step in a complete cycle, the second law may be shown to be violated unless the result is zero if the cycle is reversible, or less than zero if any step is irreversible:

$$\oint \frac{\delta q}{T} \underset{\text{irrev}}{\overset{\text{rev}}{\leqq}} 0$$

If the cyclic integral of any differential vanishes, the differential must be the exact differential of some function of the independent variables involved Since the cyclic integral of $\delta q/T$ vanishes for the reversible case, it follows that $\delta q_{rev}/T$ must be the exact differential of a function which is a property of the state of the system. This function is the entropy, designated by symbol S, and it must satisfy the relations

Reversible process $dS = \delta q/T$
Irreversible process $dS > \delta q/T$

The entropy function thus provides the basis for distinguishing between irreversible and reversible processes Although the change in entropy for a process is fixed by the initial and final states, the calculation of the change must be made for a reversible path, because only then is $dS = \delta q/T$.

For an isolated closed system (E, V constant), since δq is identically equal to zero, the criterion for equilibrium is

$$(dS)_{E,V} = 0$$

for any variation from the equilibrium state. A necessary consequence of this relation is the equivalent requirement

$$(dE)_{S,V} = 0$$

Consider now a heterogeneous system of π phases and m independent components (that is, none can be formed from others) under ordinary conditions, which implies the absence of gravitational, electric, and surface effects. For a particular differential process, the change in internal energy for phase α may be written as

$$dE^\alpha = T^\alpha\, dS^\alpha - p^\alpha\, dV^\alpha + \sum_{i=1}^{m} \mu_i^\alpha\, dn_i^\alpha \qquad (1)$$

where T^α, S^α, p^α, and V^α are the indicated variables for phase α, and n_i^α is the number of moles of component i in phase α. The quantity μ_i^α is called the chemical potential of component i for phase α, and is defined by

$$\mu_i^\alpha = \left(\frac{\partial E^\alpha}{\partial n_i^\alpha}\right)_{S^\alpha, V^\alpha, n^\alpha_{j \neq i}}$$

The chemical potential is an intensive variable whose value is determined by the relative numbers of moles of the various constituents rather than the total mass of the phase. *See* EQUILIBRIUM, PHASE.

For any differential process at equilibrium, a relation of type (1) may be written for each phase present. It is required that

$$dE = \sum_{\alpha=1}^{r} dE^\alpha = 0$$

subject to the constraints of constant total entropy, volume, and the number of moles of each independent component, that is,

$$\sum_{\alpha=1}^{r} dV^\alpha = 0 \qquad \sum_{\alpha=1}^{r} dS^\alpha = 0 \qquad \sum_{\alpha=1}^{r} dn_i^\alpha = 0$$

for all i. It may then be shown that the temperature and pressure have the same value for all phases, and that the chemical potential μ_i for any component i has the same value for every phase in which it is present. If the substances present are not all independent, but some can react chemically, [illegible] independent. It then results that the equilibrium values of the chemical potentials of the constituents involved in the reaction must be related in the sense of the stoichiometry of the reaction equation, that is, for the reaction $A \rightleftharpoons 2B$, at equilibrium $\mu_A = 2\mu_B$. Finally, while the relations specified were derived for an isolated system, they are of general validity, since if a system is at equilibrium under any conditions, the equilibrium will not be destroyed by putting a rigid adiabatic shield about the system.

Free energy. The criteria for spontaneous change, reversibility, and equilibrium for conditions of greater practical interest are readily obtained. For a closed system, the first law requires that

$$\delta q = dE + p\,dv + \delta w_{\text{net}}$$

where δw_{net} represents any work other than pv work accompanying the interaction with the surroundings. For a reversible process $T\,dS = \delta q$, and hence

$$T\,dS = dE + p\,dv + \delta w_{\text{net}} \quad \text{or} \quad dE + p\,dv - T\,dS = -\delta w_{\text{net}}$$

For the [illegible]

$$T\,dS > dE + p\,dv + \delta w_{\text{net}} \quad \text{or} \quad dE + p\,dv - T\,dS < -\delta w_{\text{net}}$$

If there is now defined a new variable

$$G = E + pv - TS = H - TS$$

the Gibbs free energy, then at constant temperature and pressure

$$dG = dE + p\,dv - T\,dS$$

and hence the criterion of reversibility becomes $dG = -\delta w_{\text{net}}$, and of irreversibility $dG < -\delta w_{\text{net}}$. For chemical equilibrium at constant T and P, since there can be no net work done, $dG = 0$, and the free energy of the system is at a minimum.

A parallel set of relations can be written for constant temperature and volume in terms of the Helmholtz free energy $A = E - TS$. The functions A and G again are extensive properties of the state of the system, and dA, dG are exact differentials.

For a single homogeneous phase, and for only pv work, there holds the relation of type (1)

$$dE = T\,dS - p\,dv + \sum_{i=1}^{m} \mu_i\,dn_i$$

For $H = E + pv$, $A = E - TS$, $G = H - TS$, there result

$$dH = T\,dS + v\,dP + \sum_{i=1}^{m} \mu_i\,dn_i \qquad (2)$$

$$dA = -S\,dT - P\,dv + \sum_{i=1}^{m} \mu_i\,dn_i \qquad (3)$$

$$dG = -S\,dT + v\,dP + \sum_{i=1}^{m} \mu_i\,dn_i \qquad (4)$$

Thus, there follow the equivalent definitions for μ_i,

$$\mu_i = \left(\frac{\partial E}{\partial n_i}\right)_{v,S,n_{j\neq i}} = \left(\frac{\partial H}{\partial n_i}\right)_{S,P,n_{j\neq i}} = \left(\frac{\partial A}{\partial n_i}\right)_{T,v,n_{j\neq i}} = \left(\frac{\partial G}{\partial n_i}\right)_{T,P,n_{j\neq i}}$$

The variation in the chemical potential μ_i with composition and other variables is expressed through the fugacity function or, more commonly, through its thermodynamic activity

$$\mu_i = \mu_i^0 + RT \ln a_i$$

Here μ_i represents the chemical potential for the state of activity a_i, and μ_i^0 the value of μ_i for a selected reference state at the same temperature. The activity concept and its application to the treatment of chemical equilibrium are considered elsewhere. *See* ACTIVITY (THERMODYNAMICS); FREE ENERGY; FUGACITY.

Since dG is an exact differential, Eq. (4) shows that

$$\left(\frac{\partial G}{\partial T}\right)_{P,\text{comp}} = -S$$

For a process involving only pv work, then

$$\left(\frac{\partial \Delta G}{\partial T}\right)_{P,\text{comp}} = \frac{\Delta G}{T} - \frac{\Delta H}{T} = -\Delta S$$

or $$\left(\frac{\partial(\Delta G/T)}{\partial T}\right)_{P,\text{comp}} = -\frac{\Delta H}{T^2}$$

This is the Gibbs-Helmholtz equation; since the thermodynamic equilibrium constant K is given by $\Delta G^0 = -RT \ln K$, where ΔG^0 is the free energy change for the reaction for standard conditions, the rate of change of K with temperature is fixed by

$$\left(\frac{\partial \ln K}{\partial T}\right)_P = \frac{\Delta H^0}{RT^2}$$

Here ΔH^0 is the standard heat of reaction.

As an example of the calculation of the effect of experimental conditions on thermodynamic properties, consider the effect of pressure on the enthalpy of a pure substance. From Eq. (2) above,

$$\left(\frac{\partial H}{\partial P}\right)_T = T\left(\frac{\partial S}{\partial P}\right)_T - V$$

Since dG is an exact differential, it follows from Eq. (4) that

$$\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P$$

and hence $$\left(\frac{\partial H}{\partial P}\right)_T = V - T\left(\frac{\partial V}{\partial T}\right)_P$$

For an ideal gas

$$V = T\left(\frac{\partial V}{\partial T}\right)_P$$

and its enthalpy is independent of pressure; for real gases at moderate pressures $(\partial H/\partial P)_T$, though not zero, is usually small, and this is why heats of reaction in the gas phase are ordinarily relatively independent of pressure.

Third law of thermodynamics. Only changes in E, H, A, and G can be measured experimentally. For the entropy, however, what are in effect absolute values can be determined from thermochemical data. Consider a pure solid such as copper, for which there is a single stable crystalline form for the range from room temperature to 0°K. Then $dS = \delta q_{rev}/T$, $q = C_p\,dT$,

$$S(T°\text{K},P) - S(0°\text{K},P) = \int_0^T \frac{C_p\,dT}{T}$$

The integral may be evaluated from specific heat data combined with an extrapolation based on the Debye theory of specific heats. (For a substance having more than one stable phase in the temperature range considered, in addition to an integration over the range of existence of each phase, there will be a contribution from each phase transition of the form $L_{\alpha\beta}/T_{\alpha\beta}$, where $L_{\alpha\beta}$ is the latent heat of transformation from phase α to β, and $T_{\alpha\beta}$ the corresponding temperature.) The third law of thermodynamics then states that $S(0°\text{K})$ may be set equal to zero for a perfect crystalline solid. (Note that entropy effects associated with isotope mixing and nuclear spin degeneracy are ignored here. See ENTROPY.) The third-law entropy $S(T,P)$ may then be calculated. Comparisons of such experimental values with those calculated by statistical thermodynamic methods have provided evidence for the validity of the third law in appropriate cases, and in others, through examination of apparent discrepancies, have lead to new conclusions concerning the structure of the solid (ice, nitric oxide), or information on the energy level system for the molecules, as in the case of restricted rotation in ethane.

The calculation of thermodynamic properties for a gas from the properties of the individual molecules is an important tool in modern thermodynamics. See STATISTICAL MECHANICS.

Thermodynamics of irreversible processes. Classical thermodynamics is primarily concerned with calculations for reversible processes, and deals with irreversibility in terms of inequalities. A quantitative treatment of irreversible processes in systems slightly removed from equilibrium is an important recent development. The rate dS'/dt of production of entropy resulting from processes within a system, as distinct from effects of interaction with the surroundings, is correlated with the fluxes J_i and generalized forces, or affinities, X_i through the relation:

$$\frac{dS'}{dt} = \sum_i J_i X_i$$

Typical forces of interest arise from mass and temperature gradients. The resulting fluxes include ordinary diffusion due to the former, and heat conduction due to the latter. In addition to these so-called direct effects, coupled effects or interactions occur; these include thermal diffusion (a mass flux due to a temperature gradient), and the Dufour effect (energy flux due to a mass gradient). The fluxes must then be related in a general way to the various affinities, and a linear dependence is assumed for systems close to equilibrium:

$$J_i = \sum_j L_{ij} X_j$$

The quantities L_{ij} are called the phenomenological coefficients. If $L_{lk} \neq 0$, there then exists a coupling between processes l and k. It has been shown by Lars Onsager that through adoption of certain definitions for the fluxes J_i and their conjugate affinities X_i, it is possible to make the set of phenomenological coefficients symmetric, that is, $L_{ij} = L_{ji}$. These conditions constitute the Onsager reciprocal relations. Proof of this proposition rests on the principle of microscopic reversibility, which requires that in a system at equilibrium, any molecular process and its reverse take place at the same average rate.

It cannot yet be clear how far-reaching the applications of this theory will be, but it provides an

approach to the treatment of the thermodynamics of the steady state, and of any process involving coupled flows. See EQUILIBRIUM, CHEMICAL.

[P.D.]

Bibliography: J. G. Aston and J. J. Fritz, *Thermodynamics and Statistical Thermodynamics*, 1959; P. W. Bridgman, *The Nature of Thermodynamics*, 1941; K. G. Denbigh, *Principles of Chemical Equilibrium*, 1955; K. G. Denbigh, *The Thermodynamics of the Steady State*, 1951; J. H. Keenan, *Thermodynamics*, 1941; I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes*, 1955

Thermoelectricity

The direct conversion of heat into electrical energy, or the reverse. By usage, the term is restricted to three processes in solid and liquid conductors: the Seebeck effect, the Peltier effect, and the Thomson effect. The last two are really detailed treatments of the first effect. These three, along with the irreversible joulean heating and conduction of heat, comprise the foundation of thermoelectricity (see JOULE'S LAW). Because the meaning of the term thermoelectricity had become fixed years before their discovery, other electrothermal phenomena, such as thermionic emission, the production of power therefrom, magnetic refrigeration, and many others, are excluded. See CRYOGENICS, THERMIONIC EMISSION.

Thermoelectric effects now have significant applications in science and engineering, and show promise of more importance in the future. These applications principally concern the measurement of temperature, generation of power, cooling, and heating. Methods for temperature measurement have been highly developed and permit great accuracy and extreme sensitivity. Applications to power generation, cooling, and heating are still in the process of development, but show increasing promise, particularly since the advent of semiconductors. A great deal of research toward the direct thermoelectric generation of electricity using the prodigal heat from nuclear reactors is underway at present. Cooling units based on thermoelectric effects have been constructed in sizes up to that of home refrigerators. Development of thermoelectric heating is not as far along as the others, but there are indications that it will have important uses, at least in small-scale heating devices

SEEBECK EFFECT

T J Seebeck discovered in 1821 that near a circuit composed of two different metals, a magnetic needle will move if the two junctions are at different temperatures.

Seebeck's experiment. In the diagram of Seebeck's experimental apparatus (Fig 1), a circuit of copper and bismuth was set with its plane vertical in the magnetic meridian; *ns* is a compass needle. With the circuit closed, when the lower Cu Bi junction was warmed, the *n* end of the needle moved toward the east. Today the interpretation of

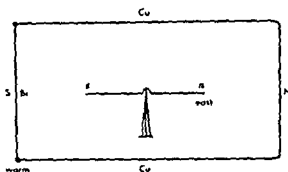


Fig 1. Diagram of apparatus used by Seebeck.

this phenomenon is that (1) there is a continuous electric current in the circuit; (2) it flows from Bi to Cu at the warmer junction and from Cu to Bi at the other (cooler) junction; (3) if the temperatures of the two junctions were interchanged, the current would be reversed; and (4) there is an emf E in the circuit. The distribution of E in the conductors will be deduced later from the Peltier and Thomson effects.

Seebeck rejected the idea that an electric current was present. He proposed that the earth's magnetism is produced by the temperature differences between equator and poles. In order to establish his views, he undertook experiments on solid and liquid metals, alloys, ores, and semiconductors The value of his work depends, not upon his hypothesis, but upon his discovery of a basic phenomenon and upon the extent and dependability of his data which are still of scientific interest.

The name thermocouple is given to the type of [illegible] two linear [illegible] nd B, the [illegible] junctions. Of paramount interest are the temperatures of the two junctions and the dependent emf E (which is variously called Seebeck emf, total thermal emf, and thermoelectric emf). The current is incidental and is derivable from E by Ohm's law. It may be of the order of microamperes or hundreds of amperes in the same thermocouple.

Thermoelectric (or Seebeck) emf. Figure 2 shows a thermocouple used in making precise measurements of E The end junctions of the conductors A and B are kept at controlled temperatures T_0 and T by immersion in or thermal coupling to baths or heat reservoirs These are indicated by the dotted rectangles. One of the conductors, B, is cut, and a measuring instrument M is inserted in the gap J_1J_2 under conditions of temperature uniformity. A potentiometer is used to measure the open circuit emf because no current through the thermocouple is required; joulean heating, with the accompanying losses to E and disturbing influences on the various temperatures involved, is avoided, and the voltage readings are independent of the resistances in the circuit. See POTENTIOMETER (VOLTAGE MEASUREMENT).

The following propositions about E in a thermocouple are the result of experimental and other

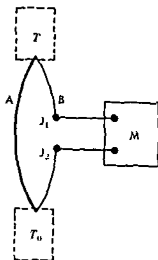


Fig. 2. Diagram of apparatus usually used for measuring thermoelectric (Seebeck) emf.

studies. Units employed in this discussion are as follows: potential and emf, in volts, millivolts (mv = 10^{-3} volt), or microvolts (μv = 10^{-6} volt); current, in amperes; electrical charge, in coulombs; and work, energy, and quantity of heat, in joules.

In a circuit kept at a uniform temperature, $E = 0$ even though it may be composed of various dissimilar conductors. This statement can be deduced by thermodynamical reasoning. If $E \neq 0$, a motor inserted in the circuit could do work, but the only source of energy would be the heat in the surroundings, which must also be at the same temperature. Thus, there would result a contradiction of the second law of thermodynamics, which states that this energy, even though present, would not be available for work.

Law of Magnus. This law, also known as the law of the homogeneous circuit, states that E depends only upon the temperatures of the two junctions and not upon the temperatures distributed (temperature gradients) along the conductors between the junctions. Apparent failures of this law have been traced, for the most part, to a lack of sufficient homogeneity in the conductors, meaning that the wires contained thermoelectrically dissimilar sections. The term homogeneity has been found experimentally to cover a state of strain as well as variation in chemical composition; an unstrained and a strained piece of wire from the same spool will show a Seebeck emf when made into a thermocouple. The presence of dissimilar sections in a thermocouple wire entails junctions whose temperatures depend upon the temperature gradient. Such potential differences are called parasitic thermo emfs.

For greater accuracy, then, the law of Magnus should specify that the conductors be homogeneous. Strictly speaking, as P. W. Bridgman has pointed out, they should also be isotropic. Experiments have shown that two rods cut from a noncubic single metal crystal in different directions may be thermoelectrically dissimilar, although highly homogeneous. Strict adherence to this condition would rule out all conductors except singl crystals, polycrystalline wires on the cubic cryst system, and liquid metals. Polycrystalline wires o noncubic metals will, however, serve very well fo thermocouples if the crystal grains are sma enough and their arrangement haphazard enoug to pass simple tests for verifying thermoelectri uniformity. The law of Magnus will be assume hereafter in the discussion. Finally, strains accom panying temperature gradients in solids and pres sure gradients in liquids undoubtedly occur, bu their thermoelectric effects are negligible. Se CRYSTAL STRUCTURE; SINGLE CRYSTAL.

Effects of junction temperatures. Values of E versus T_0 and T (Fig. 3) are reproducible to a hig degree of precision (of the order of a microvolt o less) for a given thermocouple. The relationship i empirical and cannot be predicted from theory, ex cept roughly. Neither can it be predicted from th performance of another thermocouple of sup posedly the same materials. It is determined b measuring E for a set of junction temperature which are known, such as melting points and th [illegible]

graphs, or an equation. Usually one junction temperature, T_0, is kept constant while the other, T, is varied. They are called the reference temperature and the variable or running temperature, respectively. In laboratory work, the reference temperature is usually 0°C, while in industry it is nearer to room temperature. Transformation from one system of reference temperature to the other is simple. In Fig. 3, T_0 and T are 0°C and t°C, respectively. The graphs are idealized and do not apply to any

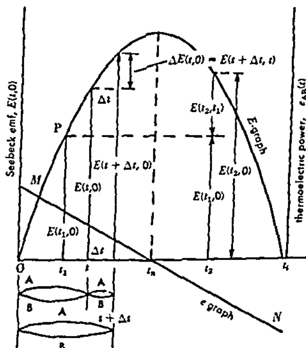


Fig. 3 Illustrative graphs of Seebeck emf and thermoelectric power. Reference junction at 0°C. E and e are written for E_{AB} and e_{AB}

particular materials, but they suffice to bring out broad features without violating any principle. A formula that often fits the data *fairly* well over ranges of 100°C or more is

$$E = at + \tfrac{1}{2}bt^2 + \tfrac{1}{3}ct^3 \qquad (1)$$

where a, b, and c are constants, and t is the temperature of one junction in degrees centigrade; the other junction is at 0°C.

Consistency demands a more detailed symbol for E, such as $E_{AB}(T,T_0)$, where A and B specify the

tion and continued around the circuit This definition of direction is commonly followed in published data, the reverse of it is more convenient

peratures 1°C and 0°C.

Table 1. Seebeck emfs for several couples

Couple	Emf, μv
E_{Fe-Cu} at 1°C, 0°C	+13 4
E_{Sb-Cu} at 1°C, 0°C	+32
E_{Bi-Cu} at 1°C, 0°C	−72 8

order of the subscripts on the left, taken with the egative sign on the right. Thermoelectric power s the name given to $E_{AB}(T + 1,T)$ or, more precisely, to the limiting value of $E_{AB}(T + \Delta T,T)/\Delta t$ as Δt approaches 0. Practically, the numerical difference is unimportant The figures given on the right in Table 1 are numerically approximately equal to the thermoelectric powers in μv/°C at 0°C. These quantities serve to compare what can be expected from various couples (Table 2).

Table 2. Thermoelectric powers at 0°C of some metallic elements against copper as reference metal; e, μv/°C

Se	Te	Ge					
+997	+397	+297					
Sb	Fe	La	Ce	Mo	Zn	Cu	
+32	+13 4	+8 7	+4 4	+3 3	+0 3	0 0	
Tl	W	Ca	Sn	Pb	Al	Pt	Hg
−0 8	−1 1	−2 4	−2 6	−2 8	−3 2	−5 9	−6 0
Na	K	Co	Ni	Bi			
−7 0	−14.3	−20 4	−20 4	−72 8			

Law of intermediate temperatures If T_0, T_1, and T_2 are temperatures in ascending order of magnitude,

$$E_{AB}(T_1,T_0) + E_{AB}(T_2,T_1) = E_{AB}(T_2,T_0) \quad (2)$$

Thus, the algebraic sum of the two Es in Fig. 4*a* equals the E of Fig 4*b*. Even if T_1 is not intermediate in magnitude, the equation holds, since the

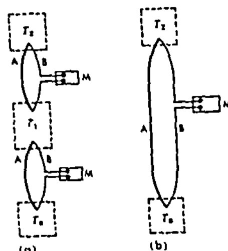


Fig 4. (*a*, *b*) Illustration of the law of intermediate temperatures; M, instruments for measuring potential.

signs will take care of themselves. Illustration of this law appears in Fig. 3 also

The law is not surprising, because it can be established by Eqs. (5) and (7) that

$$E_{AB}(T_1,T_0) = f(T_1) - f(T_0)$$
$$E_{AB}(T_2,T_1) = f(T_2) - f(T_1)$$

On adding,

$$E_{AB}(T_1,T_0) + E_{AB}(T_2,T_1) = f(T_2) - F(T_0) = E_{AB}(T_2,T_0)$$

This corresponds physically to opening the junctions at T_1 (Fig 4*a*) and uniting the conductors A to A and B to B, thus forming the single couple in Fig 4*b* The process is analogous to connecting two voltaic cells in series Thermoelectric power, being a function of temperature, may be symbolized by $e_{AB}(T)$

A convenient way of writing Eq. (2) is

$$E_{AB}(T_2,T_1) = E_{AB}(T_2,T_0) - E_{AB}(T_1,T_0) \quad (2a)$$

First, this relation shows that E for any two junction temperatures is obtained simply from data based on a fixed reference temperature by taking differences. Next, it shows that if the reference temperature is shifted (for example, from T_0 to T_1), the new values of E differ from the old by a constant, namely, the last term above. T_2 becomes the running temperature T. In Fig. 3, this shift amounts to moving the origin of coordinates to a new point P on the curve of E_{AB} versus t. Finally, returning to T_0 as the reference junction, and separating T_2 from T_1 by ΔT, Eq. (2a) may be written

$$E_{AB}(T + \Delta T,T)/\Delta T = \Delta E_{AB}(T,T_0)/\Delta T$$

which is independent of the reference temperature T_0 In the limit as $\Delta T = 0$, the equation becomes

$$e_{AB}(T) = \frac{dE_{AB}(T,T_0)}{dT}$$

or better

$$e_{AB}(T) = \left(\frac{dE}{dT}\right)_{AB}(T) \qquad (3)$$

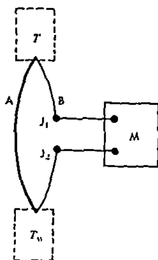


Fig. 2. Diagram of apparatus usually used for measuring thermoelectric (Seebeck) emf.

studies. Units employed in this discussion are as follows: potential and emf, in volts, millivolts (mv = 10^{-3} volt), or microvolts (μv = 10^{-6} volt); current, in amperes; electrical charge, in coulombs; and work, energy, and quantity of heat, in joules.

In a circuit kept at a uniform temperature, $E = 0$ even though it may be composed of various dissimilar conductors This statement can be deduced by thermodynamical reasoning. If $E \neq 0$, a motor inserted in the circuit could do work, but the only source of energy would be the heat in the surroundings, which must also be at the same temperature. Thus, there would result a contradiction of the second law of thermodynamics, which states that this energy, even though present, would not be available for work.

Law of Magnus This law, also known as the law of the homogeneous circuit, states that E depends only upon the temperatures of the two junctions and not upon the temperatures distributed (temperature gradients) along the conductors between the junctions. Apparent failures of this law have been traced, for the most part, to a lack of sufficient homogeneity in the conductors, meaning that the wires contained thermoelectrically dissimilar sections. The term homogeneity has been found experimentally to cover a state of strain as well as variation in chemical composition; an unstrained and a strained piece of wire from the same spool will show a Seebeck emf when made into a thermocouple. The presence of dissimilar sections in a thermocouple wire entails junctions whose temperatures depend upon the temperature gradient. Such potential differences are called parasitic thermo emfs.

For greater accuracy, then, the law of Magnus should specify that the conductors be homogeneous. Strictly speaking, as P. W. Bridgman has pointed out, they should also be isotropic. Experiments have shown that two rods cut from a noncubic single metal crystal in different directions may be thermoelectrically dissimilar, although highly homogeneous. Strict adherence to this condition would rule out all conductors except singl crystals, polycrystalline wires on the cubic crysta system, and liquid metals. Polycrystalline wires o noncubic metals will, however, serve very well fo thermocouples if the crystal grains are smal enough and their arrangement haphazard enougl to pass simple tests for verifying thermoelectri uniformity. The law of Magnus will be assume hereafter in the discussion. Finally, strains accompanying temperature gradients in solids and pressure gradients in liquids undoubtedly occur, bu their thermoelectric effects are negligible. *See* CRYSTAL STRUCTURE; SINGLE CRYSTAL.

Effects of junction temperatures. Values of E versus T_0 and T (Fig. 3) are reproducible to a high degree of precision (of the order of a microvolt or less) for a given thermocouple. The relationship is empirical and cannot be predicted from theory, except roughly. Neither can it be predicted from the performance of another thermocouple of supposedly the same materials. It is determined by measuring E for a set of junction temperatures which are known, such as melting points and the

graphs, or an equation. Usually one junction temperature, T_0, is kept constant while the other, T, is varied. They are called the reference temperature and the variable or running temperature, respectively. In laboratory work, the reference temperature is usually 0°C, while in industry it is nearer to room temperature. Transformation from one system of reference temperature to the other is simple. In Fig 3, T_0 and T are 0°C and t°C, respectively. The graphs are idealized and do not apply to any

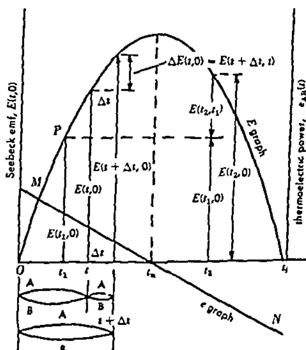


Fig. 3. Illustrative graphs of Seebeck emf and thermoelectric power. Reference junction at 0°C. E and e are written for E_{AB} and e_{AB}.

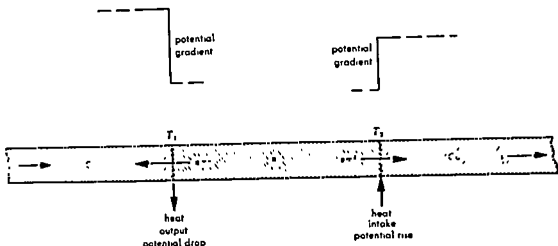


Fig 6. Diagram illustrating Peltier effect. Small current i directed to right

the Peltier rate of heat transfer is proportional to $F_i i$, that is, proportional to i. However, i may be taken as small as desired so that the joulean heat can be neglected. In this discussion, the simplified description of the current in terms of moving positive charges which produce the same magnetic field as the actual moving charges, whatever their signs, is followed. Also, energy relations are worked out by supposing 1 coulomb of positive charge to move slowly around the circuit in the direction of the ac- [illegible]

[illegible] discovery of Icilius that, for every coulomb passing a given junction, the amount of heat intake or output is the same for all values of the current. The only variables left to affect heat intake or output are the materials and temperature. Suppose 1 coulomb of positive charge passes slowly along the conductor of Fig. 6. At the first junction, the heat output in joules must equal the decrease of the potential energy of the charge, internal energy changes and external work being eliminated. This decrease is equal numerically to a potential drop in volts at this junction. In a similar manner, there is a potential rise at the second junction in volts equal to the heat intake per coulomb Both potential changes are also functions of materials and temperatures only. The potential gradient (Fig 6) illustrates this; it is independent of the current whether direct, reverse, or zero Therefore, the first junction is the seat of an opposing emf and the second, of an assisting emf; both emfs are directed from Bi to Cu. By a commonly accepted notation, the heat intake in joules when 1 coulomb of positive charge passes from conductors A to B at a temperature T is written $\Pi_{AB}(T)$ and is variously called the Peltier heat, the Peltier emf, or the Peltier coefficient. Thus, at the first junction in Fig 6, $\Pi_{CuBi}(T_1)$ would be written, having negative numerical values; or $-\Pi_{BiCu}(T_1)$, having positive values At the second junction, $\Pi_{BiCu}(T_2)$ would have positive values. The Peltier emf should not be confused with the contact potential or Volta effect, which is the potential difference developed between two dissimilar metals when touched and separated. Seebeck is credited with demonstrating the difference between the Volta effect and the effect on a magnetic needle produced by temperature differences. This conclusion was a by-product of his extensive researches.

Thermocouples. In the thermocouple of Fig. 1, the direction of the current i is counterclockwise. The temperatures of the warm and cool junctions are assumed to be T and T_0, respectively. The two Peltier emfs are directed from Bi to Cu at each junction irrespective of the source of the current i. Their resultant in the direction of i is their difference. If there were no other emf in the circuit,

$$\Pi_{BiCu}(T) - \Pi_{BiCu}(T_0) = E$$

the Seebeck emf However, Lord Kelvin showed by thermodynamical reasoning that this equation leads to a conflict with experience. If 1 coulomb of positive charge moves slowly around the circuit, there will be Peltier cooling in its passage from Bi to Cu at T and Peltier heating at T_0. To maintain the Ts constant during this cycle, heat must be supplied at T and rejected at T_0. This amounts to transferring the emphasis from emfs to quantities of heat in the above equation, because the first and second terms also represent these heats, respectively The third term represents the energy available for work. Thus, assuming that the heat transfers and performance of work may be effected reversibly, the thermocouple may be treated as an ideal heat engine for which it has been shown that the ratio (heat intake)/(heat output) = T/T_0 where the Ts are absolute thermodynamic temperatures Therefore, in abbreviated symbols

$$\Pi - \Pi_0 = E \quad \text{and} \quad \Pi/\Pi_0 = T/T_0$$

may be written, whence it follows that

$$(\Pi - \Pi_0)/\Pi_0 = (T - T_0)/T_0$$

or

$$E = (\Pi_0/T_0)(T - T_0)$$

Now if T_0 is kept constant, as in Fig. 3, E would be an increasing linear function of T, which is not supported by experiment. There could be no neu-

since the terms are independent of T_0 Here the left-hand member is the thermoelectric power while the derivative on the right is represented by the slope of the E-T curve, Fig. 3, for any reference temperature. As an illustration, Eq. (1) yields $e_{AB} = a + bt + ct^2$ or, if c is neglected, $e_{AB} = a + bt$ (as is assumed for Fig 3). The plot of this is the straight line MN, which crosses the t axis at t_n, called the neutral temperature, and takes on negative values beyond. The temperature at which E reverses, t_i, is called the inversion temperature; $t_n = -a/b$, and t_i is given by the solution of

$$at + \tfrac{1}{2} bt^2 = 0$$

The results would be expressed in degrees centigrade. Graphs of actual values of e, especially for temperature ranges of the order of 1000°C, are far from straight lines, but no new principle is involved. The values of e given in Table 2 are arranged in descending order. The order may be different at temperatures other than 0°C. Iron and copper, for example, reverse their order above t_n, or about 500°C

Law of intermediate metals. The second of the two additive relations states that

$$E_{AB}(T,T_0) + E_{BC}(T,T_0) = E_{AC}(T,T_0) \quad (4)$$

It is evident that, if the two couples AB and BC in Fig. 5*a* are merged, all emfs in the metal B will cancel so that B may be removed, and the over-all emf becomes $E_{AB}(T,T_0)$. An important corollary is illustrated in Fig 5*b*, where a third metal B is inserted in a thermocouple AC as shown to make junctions at equal temperatures. E_{AC} is not changed thereby. Thus the use of solder or welding at junctions will not alter E

For similar reasons (Fig. 2), the readings of M will not be in error provided that the temperatures of the junctions J_1 and J_2 are equal and that J_1MJ_2 is composed of homogeneous material, or if it is not homogeneous, that the whole of the material also is at the same temperature The junctions J_1 and J_2 are kept at the same temperatures by immersion in a bath, which often is that of T_0 In order that Table 2 may be used generally, Eq (4) may be written

$$E_{AB}(T,T_0) = E_{AC}(T,T_0) - E_{BC}(T,T_0) \quad (4a)$$

By differentiation,

$$e_{AB}(T) = e_{AC}(T) - e_{BC}(T)$$

Application to Sb and Bi in Table 2 yields

$$e_{SbBi} = e_{SbCu} - e_{BiCu} = 32 + 72.8 = 104\,8\ \mu v/°C$$

at 0°C, which shows how to use such tables for comparing various pairs of metals A similar rule holds, of course, for the Es. The further apart two metals are found in the series, the more powerful is the couple formed from them. Seebeck constructed an extensive and still useful thermoelectric series whose order agrees very well with present-day findings.

Two phenomena, already mentioned, remain to be considered. One, the Peltier effect, interpreted

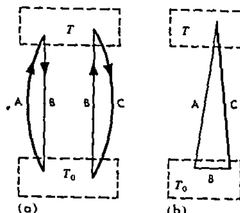


Fig. 5. *(a, b) Illustration of the law of intermediate metals.*

according to the energy conservation law, serves to localize a part of the Seebeck emf at the junctions. Similarly the second, the Thomson effect, serves to find the remainder distributed along the two conductors in terms of the temperature gradients present in them The two effects taken together are essential to the proofs relating thermoelectricity to the mechanical theory of heat, that is, to thermodynamics. This step stands out as one of the landmarks in the development of the physical sciences. *See* THERMODYNAMICS (CHEMICAL).

PELTIER EFFECT

In 1834, J. C. A Peltier discovered the phenomenon which is chiefly the reverse of the Seebeck effect, namely, that cooling or heating of the junction of two dissimilar metals occurs when an electric current passes through them Experiments were made with conductors containing sections of many different metals and carrying weak currents. Where copper and bismuth (among the first metals investigated) were joined, a temperature rise was observed when the current was directed from Cu to Bi and a fall when the current was reversed. Both cases are illustrated in Fig 6 at the first and second junctions, respectively, counting from left to right. The respective junction temperatures are T_1 and T_2, of which T_1 is the higher Peltier established this, first with separate thermocouples, and later with volume changes of a gas in bulbs surrounding the junctions Obviously there is an outflow of heat at the first junction and an inflow at the second, because the temperature of the room would be between T_1 and T_2 T_1 and T_2 would reach equilibrium values finally, and the processes would become steady. A second experimental fact is important in this connection Quintus Icilius (1853) showed that the rate of heat output or intake is proportional to the current i. A third pertinent experimental fact is that no physical or chemical change has ever been detected in the conductors while a steady current is flowing through them and the temperature is constant. Therefore, the internal energy of the conductors does not change.

Theory. Theoretical discussion of the Peltier effect is simplified if the current is assumed to be small. The ratio of the rate of joulean heating to

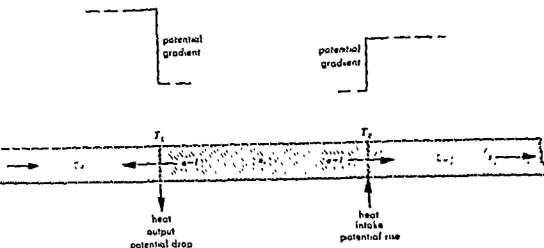


Fig 6. Diagram illustrating Peltier effect Small current i directed to right

the Peltier rate of heat transfer is proportional to $F_1 i$, that is, proportional to i. However, i may be taken as small as desired so that the joulean heat can be neglected. In this discussion, the simplified description of the current in terms of moving positive charges which produce the same magnetic field as the actual moving charges, whatever their signs, is followed. Also, energy relations are worked out by supposing 1 coulomb of positive charge to move slowly around the circuit in the direction of the actual current. The change in its energy in joules

a given junction, the amount of heat intake or output is the same for all values of the current. The only variables left to affect heat intake or output are the materials and temperature. Suppose 1 coulomb of positive charge passes slowly along the conductor of Fig. 6 At the first junction, the heat output in joules must equal the decrease of the potential energy of the charge, internal energy changes and external work being eliminated This decrease is equal numerically to a potential drop in volts at this junction In a similar manner, there is a potential rise at the second junction in volts equal to the heat intake per coulomb. Both potential changes are also functions of materials and temperatures only. The potential gradient (Fig. 6) illustrates this, it is independent of the current whether direct, reverse, or zero Therefore, the first junction is the seat of an opposing emf and the sec-

ture T is written $\Pi_{AB}(T)$ and is variously called the Peltier heat, the Peltier emf, or the Peltier coefficient. Thus, at the first junction in Fig 6, $\Pi_{CuBi}(T_1)$ would be written, having negative numerical values, or $-\Pi_{BiCu}(T_1)$, having positive values. At the second junction, $\Pi_{BiCu}(T_2)$ would have positive values The Peltier emf should not be confused with the contact potential or Volta effect, which is the potential difference developed between two dissimilar metals when touched and separated. Seebeck is credited with demonstrating the difference between the Volta effect and the effect on a magnetic needle produced by temperature differences. This conclusion was a by-product of his extensive researches

Thermocouples. In the thermocouple of Fig 1, the direction of the current i is counterclockwise. The temperatures of the warm and cool junctions are assumed to be T and T_0, respectively The two Peltier emfs are directed from Bi to Cu at each junction irrespective of the source of the current i. Their resultant in the direction of i is their difference If there were no other emf in the circuit,

$$\Pi_{BiCu}(T) - \Pi_{BiCu}(T_0) = E$$

the Seebeck emf However, Lord Kelvin showed by thermodynamical reasoning that this equation leads to a conflict with experience If 1 coulomb of positive charge moves slowly around the circuit, there will be Peltier cooling in its passage from Bi to Cu at T and Peltier heating at T_0 To maintain the Ts constant during this cycle, heat must be supplied at T and rejected at T_0. This amounts to transferring the emphasis from emfs to quantities of heat in the above equation because the first and second terms also represent these heats, respectively The third term represents the energy available for work. Thus, assuming that the heat transfers and performance of work may be effected reversibly, the thermocouple may be treated as an ideal heat engine for which it has been shown that the ratio (heat intake)/(heat output) $= T/T_0$ where the Ts are absolute thermodynamic temperatures Therefore, in abbreviated symbols

$$\Pi - \Pi_0 = E \quad \text{and} \quad \Pi/\Pi_0 = T/T_0$$

may be written, whence it follows that

$$(\Pi - \Pi_0)/\Pi_0 = (T - T_0)/T_0$$

or

$$E = (\Pi_0/T_0)(T - T_0)$$

Now if T_0 is kept constant, as in Fig 3, E would be an increasing linear function of T, which is not supported by experiment There could be no neu-

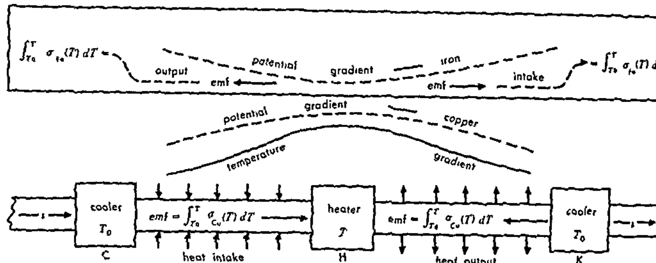


Fig. 7. Thomson effect. Lower part refers to copper, upper to iron; the two include all types of variation.

tral or inversion temperatures. Therefore, the supposition that the Peltier emfs alone are responsible for E is untenable. W. Thomson sought for additional emfs in the body of the conductors, where he *found them* associated with temperature gradients. It is obvious that the above reasoning is not limited to bismuth and copper.

THOMSON EFFECT

The third thermoelectric effect, discovered in 1854 by William Thomson, later Lord Kelvin, is essentially a reversible heat transfer between a homogeneous conductor carrying a current and its surroundings, conditioned upon the presence of a temperature gradient. Figure 7 is a schematic diagram of an apparatus used by Thomson to demonstrate the effect

Theory. If the two junctions of a thermocouple are located near the midpoints of the sections CH and HK to make $E = 0$ before the current is turned on, then after the current starts, the thermocouple shows a finite E *indicating a higher temperature in* the "downstream" section HK (in the case of copper). The temperature gradient is distorted, its maximum being shifted with the current This is true also if the current is reversed. To prevent this distortion, heat must be supplied to the CH section and abstracted from the HK section, the amounts being properly distributed. If 1 coulomb of positive electricity is supposed to pass slowly along the conductor, this distribution may be expressed by $\sigma_{Cu}(T)\,dT$. It is called a Thomson heat and is defined as the heat intake (in joules) by any section at temperatures T to $T + dT$ when 1 coulomb (+) passes through it up the potential gradient. In the case of copper, it is up the temperature gradient also, that is, $\sigma_{Cu}(T)$ is positive. It is also a uniform, finite, and continuous function of T. It follows that *the section is the seat of an emf equal to* $\sigma_{Cu}(T)\,dT$ directed up the potential and temperature gradients. The case for iron (also investigated by Thomson) *would be expressed by* $\sigma_{Fe}(T)\,dT$, the only difference being that σ_{Fe} would have negative values. The emf would still be directed up the potential gradient, but *down the temperature* gradient. It would have a minimum over the heater, and *its shift would* be against the starting current. The Thomson coefficients $\sigma_{Cu}(T)$ and $\sigma_{Fe}(T)$ reverse their signs at temperatures well below those of the Thomson experiments, that is, 130°K and 170°K respectively.

The expression for the Seebeck emf taken as the resultant of both the Peltier and Thomson effects now becomes (Fig. 8)

$$E_{AB} = \Pi_{AB}(T) - \Pi_{AB}(T_0) + \int_{T_0}^{T} (\sigma_A - \sigma_B)\,dT \quad (5)$$

where the positive direction around the circuit is taken clockwise, that is, from A to B *at the warmer* junction. The differential form of the above would be

$$\frac{dE_{AB}}{dT} = \frac{d\Pi_{AB}}{dt} + (\sigma_A - \sigma_B) \quad (6)$$

obtained either by setting the limits at T and $T + dT$ or by a differentiation with T_0 constant. The basis of Eqs. (5) and (6) is the first law of thermodynamics, and according to Kelvin involves no hypothesis Equation (5) does not contradict the law of Magnus, because the value of the integral is a function of the limits only Also, *being in the form*

$$\int_{T_0}^{T} (\sigma_A - \sigma_B)\,dT = f(T) - f(T_0) \quad (7)$$

it has proved useful in establishing the law of intermediate temperatures, Eq. (2)

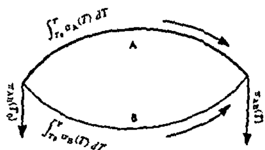


Fig. 8. Combination of Peltier and Thomson emfs in a thermocouple.

Thermodynamic relations. There remains Thomson's application of the second law of thermodynamics. Any derivation of an equation from this law would require conditions of reversibility. Heat conduction and joulean heat must be reckoned with because they are irreversible. No ideal experiment can be designed, however, which will render such processes negligible simultaneously. Thomson recognized the importance of this fact, but being convinced that the mechanisms of these two processes were independent of those creating the thermoelectric emfs, he hypothesized that they could be ignored. The derivation of an additional relation then becomes brief. Figure 9 represents a thermally isolated system composed of a *thermocouple and* four heat reservoirs at the absolute thermodynamic temperatures indicated. All heat transfers are re-

passes slowly *around the circuit in the clockwise* direction, the entropy of the couple remains unchanged because its constitution has remained unchanged. The total entropy loss of the reservoirs is

$$\frac{\Pi_{AB}(T+\Delta T)}{T+\Delta T}+\frac{\Pi_{BA}(T)}{T}+\frac{\sigma_A\,\Delta T}{T+(\Delta T/2)} + \frac{\sigma_B(-\Delta T)}{T+(\Delta T/2)} = 0$$

by the second law. The combination of the first two terms is

$$\frac{d}{dT}\left(\frac{\Pi_{AB}}{T}\right)\Delta T$$

Therefore
$$\frac{d}{dT}\left(\frac{\Pi_{AB}}{T}\right)+\frac{\sigma_A-\sigma_B}{T}=0 \qquad (8)$$

A combination of Eqs. (6) and (8) gives

$$\frac{\Pi_{AB}}{T}=\frac{dE_{AB}}{dT} \qquad (9)$$

$$-\frac{\sigma_A-\sigma_B}{T}=\frac{d^2E_{AB}}{dT^2} \qquad (10)$$

which are the celebrated Kelvin relations. Much experimental research has been done to check the theory. The right-hand members involve electrical measurements only, which are capable of high pre-

not uncommon. No experimental evidence invalidating the relations has been found; yet, dissatisfaction with Thomson's hypothesis prompted many disputes and many attempts to get around the difficulty. A theory appropriately called thermodynamics of irreversible processes has arisen, based largely on a theorem of L. Onsager (1931) and certain concepts of entropy flow and entropy production. This theory is considered satisfactorily rigorous. Its treatment is too involved with other fields to be appropriate here. See THERMODYNAMICS (CHEMICAL). The facts of interest are the deduction of Thomson's hypothesis from it (showing him to be the first to enter this field) and that the Kelvin relations still stand.

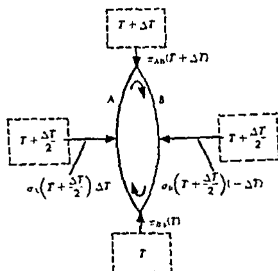


Fig. 9. Applications of second law of thermodynamics under Thomson's hypothesis. The Π and σ expressions represent heat transfers.

It should be emphasized that the quantities E, e, Π, and σ are profoundly affected by anything that alters the structure of the material of the conductors, such as impurities, heat-treatment, drawing, *rolling, pressure, tension, and magnetization.* Any attempt to correlate the Peltier and Thomson effects with other properties of matter would come under solid-state physics.

APPLICATIONS

Measurement of temperature. From measurement of E, made with a thermocouple, and the empirically determined tables, curves, or formulas discussed under Seebeck effect, temperature values can be determined very accurately and precisely. This is one of the two most important methods in precision thermometry, the other being based on electrical resistance. The sensitivity of the thermo*couple technique is startling; detection of radia*tion from a candle 50 miles distant is within its *reach. The useful range of a thermocouple is lim*ited by insensitivity at low temperatures and instability at high. A battery of thermocouples connected in series (Fig. 10) is called a thermopile. Thermocouples of copper vs a copper-nickel alloy called Constantan are widely used in the range from −169°C (104°K) up to 386°C, but are completely useless below about 15°K. Those of platinum vs an alloy of platinum and 10% rhodium can be used from near 0 to 1710°C. These two types represent the limits of the method. Superconductors show no *thermoelectric effect. Among* the many advantages of thermocouples are their rug-

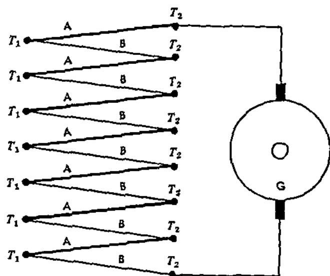


Fig. 10. Thermopile, a battery of couples connected in series.

gedness, response, applicability to inaccessible locations, simplicity and availability of materials, and adaptability to automatic measuring, recording, and control of temperatures. Couples are important assets in steel, metals, oil, and other industries. Well over 200 tons of thermocouple materials is supplied to industry annually in the United States. Thermocouples are also valuable tools of scientific research. They are often component parts of instruments which measure other physical quantities, such as radiometers, vacuum gages, and hot-wire ammeters For many decades, measurement of temperature was the only successful application of thermoelectric effects; today it may be considered practically a completely developed art. *See* TEMPERATURE MEASUREMENT; THERMOCOUPLE.

Generation of power. Obtaining power directly from heat was one of the first applications of thermoelectricity attempted, but it soon was abandoned because of the low efficiencies of the conversion Some small units were constructed, however, at a fairly early time: Peltier used a thermocouple as a source of current in some of his research, and G S Ohm also used the thermocouple in the work leading to the discovery of his famous law (1826) Later, thermopiles were used in telegraphy and electroplating

With the advent of the semiconductor and with applications of solid-state theory, interest in thermoelectric effects as a source of power has revived Experimentation in utilizing the heat from nuclear reactors as a direct source of electric power is being pursued vigorously in the Soviet Union, the United States, and in laboratories all over the world.

Future development of thermoelectric generation of power is unlikely, except for small units. The conventional methods now in use probably will not become obsolete. The same situation exists to a lesser degree for refrigeration and warming.

Pure and applied research in thermoelectricity at present centers around a quest for new thermoelectric materials better suited to power generation, heating, and cooling. Theory has indicated a guide for the researcher which is applicable in these fields. It is a "figure of merit," e^2c/k, where e = thermoelectric power, c = electrical conductivity, and k = thermal conductivity. Thus, e and c should be large, and k should be small. Stability at high temperatures is also desirable for high efficiencies in power generation. Again, metals suffer from limitations set on the value of the ratio c/k (Wiedemann and Franz law), while semiconductors offer a freer range of values.

Metals give values for e around 20–50 $\mu v/°C$ while semiconductors may show 1000 $\mu v/°C$. For the over-all efficiency of a thermoelectric generator made of metals, the maximum that can be expected ranges from 0.1 to 0.6%, although for Bi-Sb, 3.1% has been calculated. For a number of semiconductors, 7% has been reported in the literature; 10% has been approached; and 10–15% and more cannot be regarded as unattainable; 20–30% is visualized for ceramics because of their ability to withstand higher temperatures than semiconductors.

Figure 10 serves as a diagram for all these types. On the supposition that $T_2 > T_1$ and that G is a motor, heat must be supplied to the junctions at T_2 and abstracted from the junctions at T_1, while G delivers mechanical energy If G is a generator that reverses the current, then T_2 remains higher than T_1 because of the Peltier effects. Heat is absorbed at T_1 and rejected at T_2 after the heat equivalent of the power input from the generator has been added to it. The T_1 junctions may serve to refrigerate; the T_2 junctions, to warm. Any reversal of the current always interchanges the locations of these two functions. Thermoelectricity furnishes an excellent realization of the so-called heat pump, for which Kelvin pointed out the economy for heating about 1852. *See* HEAT PUMP

Thermoelectric generators. Some examples of thermoelectric generators include (1) a unit operated by a kerosine lamp capable of delivering 3 watts of electrical power at 150 volts, manufactured for use in remote regions with no electricity, (2) similar kerosine generators delivering 15–20 watts for use in some transmitters and in tractors; (3) an experimental 100-watt unit utilizing solar energy; (4) a 200-watt generator consuming 2 kg of firewood per hour, and (5) 200–500-watt models, in course of manufacture in the Soviet Union. A demonstration model powered by a gas flame was used to operate a 10-watt public address system for a lecture delivered by C Zener before a conference on thermoelectricity in Washington, D C, September, 1958. No large-scale units (above 1 kw) have been reported so far.

Thermoelectric cooling. Units the size of home refrigerators and smaller have been constructed and can compete with the compressor type, particularly the smaller sizes where it is more important to avoid weight or complex machinery than to save power, and where the compressor type of cooling device is impractical. The ratio of heat abstracted from the cooling chamber to energy input (coeffi-

cient of performance), for a temperature difference of 40°C and a value 0.4–0.7, seems practical. In the case of heating, the corresponding practical coefficient of performance (heat delivered/energy input) is 1.5–1.8 for the same temperature difference.

An example of a cooling unit is a home refrigerator of 55 liters capacity for the cooled chamber. The material of the thermopile is a PbTe-PbSe alloy against an alloy whose principal constituents are tellurium and antimony. The average maximum difference in thermoelements is 47°C; refrigerating capacity, 20 kcal/hr, dc power input 40 watts. The temperature of the chamber is −2°C when the outside is at 19°C. The cooling time is 4–5

their heat to the surroundings through fins. Some refrigerators are water-cooled. In one of these, the total weight of semiconductor materials used in its cooling units is around 50 g and its energy consumption is approximately 50 watts. In the field of low-temperature cooling, proper choice of materials can result in a reduction of temperature in the vicinity of the junctions by about 90°C.

Various devices which are dependent upon small cooling units have been developed: a dew-point meter using a silvered surface (its lightness and rapid action fit it for use in an airplane); a cold trap for vacuum systems; thermostats capable of operating above or below room temperature, including one at 0°C, used for reference junctions, controlled by the formation of ice to cut off the cooling; a cooling device for microtomes which holds the tissue at an optimum cutting temperature and can cut sections 2 microns thick.

Thermoelectric heating. This, in principle, is a refrigerator with the current reversed. No description of a large installation is available. There has

a heating cycle to warm the bottle. The thermoelectric effect promises countless possibilities for home use.

Definite information on the composition of semiconductors is not available. They seem satisfactory for use at the mild temperatures involved in refrigeration and warming, but they have not proved satisfactory at the higher temperatures appropriate to power generation. Therefore, ceramics mixed valence compounds of the transition metals have been developed for use in this temperature range. They are chemically inert and stable at the temperatures to be encountered. In view of other advantages also, they are considered promising for use at temperatures of 3000°F. *See* ELECTRICITY

[L C H]

Bibliography. Am Inst Physics, *Temperature, Its Measurement and Control in Science and Industry*, vol. 1, 1941; P. W. Bridgman, *The Thermodynamics of Electrical Phenomena in Metals*, 1934; S. R. deGroot, *Thermodynamics of Irreversible Processes*, 1951; K. G. Denbigh, *The Thermodynamics of the Steady State*, 1951; P. H. Dike, *Thermoelectric Thermometry*, 1954; W. E. Forsythe (ed.), *Smithsonian Institution Physical Tables*, 1954; *International Critical Tables of Numerical Data, Physics, Chemistry, and Technology*, vol. 6, 1929; A. F. Joffe, The revival of thermoelectricity, *Sci. American*, 199(5):31–37, 1958; A. F. Joffe, *Semiconductor Thermoelements and Thermoelectric Cooling*, 1957; M. W. Zemansky, *Heat and Thermodynamics*, 2d ed., 1943, 4th ed., 1957.

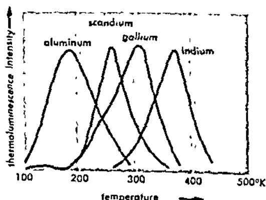
Thermoluminescence

A term sometimes used broadly to mean any luminescence appearing in a material due to the application of heat. More frequently the term refers specifically to the luminescence appearing as the temperature of the material is steadily increased.

Many solids that contain luminescent centers often also contain one or more types of centers that trap electrons. If the solid is exposed to light of sufficiently short wavelength or to x rays or other high-energy radiation, free electrons are produced in the solid and some of these electrons may be trapped. If the depth of the trap (that is, the amount of energy required to release the electron from the trap) is large and the temperature is low, the electron will remain trapped for a long time. If, however, the temperature of the sample is raised slowly, the electron will receive increasing amounts of thermal energy and will eventually escape the trap. An electron thus freed from a trap may go over to a luminescent center, give its energy to the center, and cause it to luminesce. For a single type of trap the glow curve (plot of thermoluminescence intensity as a function of temperature) first rises, reaches a maximum, and then decreases to zero when all the traps become emptied. The depth of the trap, E (in ergs), is given to a good approximation by

$$E = 1.51kT^*T'/(T^* - T')$$

where k is Boltzmann's constant, T^* is the temper-



Glow curves for several zinc sulfide phosphors, each of which contains traces of copper and different trivalent ions. The luminescent center is due to the presence of copper, the activator in each case. The traps are put in by various trivalent coactivators as shown.

ature (in degrees Kelvin) of the solid at the peak of the curve, and T' is the temperature on the low-temperature side at which the emission is one-half its peak value. The figure shows glow curves for several zinc sulfide phosphors. The incorporation of impurities (activators, coactivators) into "host" compounds such as zinc sulfide is one of the most important processes for preparing solid materials; for a detailed discussion *see* LUMINESCENCE. When more than one type of trap is present, the glow curve consists of a corresponding number of peaks, which often may be resolved and analyzed as described. Thus thermoluminescence may be used to secure information about the properties of traps in solids. *See* TRAPS IN SOLIDS. [C.C.K.; J.H.S.]

Thermomagnetic effects

Electrical and thermal phenomena occurring when a conductor or semiconductor which is carrying a thermal current (that is, is in a temperature gradient) is placed in a magnetic field.

Let the temperature gradient be transverse to the magnetic field H_z, for example along x. Then the following transverse-transverse effects are observed:

1. Ettingshausen-Nernst effect, an electric field along y,

$$E_y = Q(\partial T/\partial x)H_z$$

where Q is known as the Ettingshausen-Nernst coefficient. This coefficient is related to the Ettingshausen coefficient P by

$$P = QT\sigma$$

where σ is the thermal conductivity in a transverse magnetic field This relation was discovered by P. W. Bridgman; it has been shown to be an example of the Onsager reciprocity relations of irreversible thermodynamics. *See* GALVANOMAGNETIC EFFECTS; THERMODYNAMICS (CHEMICAL); THERMOELECTRICITY.

2. Righi-Leduc effect, a temperature gradient along y,

$$(\partial T/\partial y) = S(\partial T/\partial x)H_z$$

where S is known as the Righi-Leduc coefficient.

Also, the following transverse-longitudinal effects are observed:

3. An electric potential change along x, amounting to a change of thermoelectric power

4. A temperature gradient change along x, amounting to a change of thermal resistance.

Let the temperature gradient be along H Then changes in thermoelectric power and in thermal conductivity are observed in the direction of H.

For related phenomena *see* HALL EFFECT; MAGNETORESISTANCE. [E.A.; F.K.]

Bibliography: *See* GALVANOMAGNETIC EFFECTS.

Thermometer

An instrument that measures temperature. Although this broad definition includes all temperature-measuring devices, they are not all called thermometers. Other names have been generally adopted. For discussion of two such devices, *see* PYROMETER; THERMOCOUPLE. For general discussion of temperature measurement, *see* TEMPERATURE MEASUREMENT.

A variety of techniques are used in instruments known as thermometers. Some of these depend on the expansion of a liquid or metal for the indicating means. Others employ the change in pressure of a gas to detect the temperature. Still others use the change in electrical resistance which occurs with temperature changes.

Liquid-in-glass thermometer. This thermometer consists of a liquid-filled glass bulb and a connecting partially-filled capillary tube. When the temperature of the thermometer increases, the differential expansion between the glass and the liquid causes the liquid to rise in the capillary. In Fig. 1*a* the graduations are etched on the glass stem. The thermometer in Fig. 1*b* has a separate graduated scale similar to that of the common household thermometer. A variety of liquids, such as mercury, alcohol, and pentane, and a number of different glasses are used in thermometer construction, so that various designs cover diverse ranges between about −300°F and +1200°F. Expansion and contraction chambers are sometimes provided at each end of the capillary to permit over-range and un-

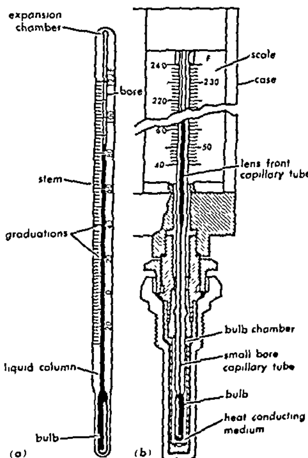


Fig. 1 Liquid-in-glass thermometers. (a) Etched-stem clinical thermometer. (b) Graduated-scale industrial thermometer. (From D. M Considine, ed, Process Instruments and Controls Handbook, McGraw-Hill, 1957)

der range use of the thermometer without loss of accuracy When the entire thermometer is not subjected to the same temperature, an error occurs unless the thermometer is calibrated for these conditions Many thermometers are used with an emergent stem and are calibrated for this type of service.

Maximum-registration thermometers, such as a lever thermometer, allow expansion of the liquid with increasing temperature but maintain the liquid column as the temperature decreases, thereby indicating the maximum value attained. Minimum registering thermometers are also available. If the liquid filling is metallic, electric contacts can be mounted in the stem wall to complete a circuit when temperature reaches a specified value. These are used in alarm and control systems. Some thermometers destined for rugged service are enclosed in a metal sleeve or armor.

Accuracies and speeds of response vary widely with the designs, ranges, and installations. For example, the Beckmann thermometer can be easily read to an accuracy (stated in terms of maximum error) of 0.001°C and the fever thermometer to 0.1°F Industrial thermometers and armored thermometers are seldom more accurate than 1°F as they are used. Mercury and glass thermometers have time constants as low as 0.1 second in well-stirred water, but industrial thermometers and all

[illegible]

differential expansion of thin dissimilar metals, bonded together into a narrow strip and coiled into the shape of a helix or spiral, is used to actuate a pointer (Fig. 2) Case designs are available for laboratory or heavy industrial service Range spans are seldom shorter than 50°F or longer than 400°F, with a maximum upper temperature limit for continuous service of 800°F and a minimum of −300°F The shorter range spans are used near room temperatures, and accuracies in the neighborhood of 1°F can be achieved At high and low

Fig. 2 Bimetal thermometer. (Weston Instruments, Division of Daystrom, Inc)

Fig. 3 Filled-system thermometer. (*D. M Considine, ed., Process Instruments and Controls Handbook, McGraw-Hill, 1957*)

temperatures the accuracy is seldom better than 5°F The time constants of these thermometers are greater but of the same order of magnitude as the liquid-in-glass thermometer.

Filled-system thermometer. This type of thermometer, shown schematically in Fig. 3, has a bourdon tube connected by a capillary tube to a hollow bulb. When the system is designed for and filled with a gas (usually nitrogen or helium) the pressure in the system substantially follows the gas law, and a temperature indication is obtained from the bourdon tube The temperature-pressure motion relationship is nearly linear. Atmospheric pressure effects are minimized by filling the system to a high pressure When the system is designed for and filled with a liquid, the volume change of the liquid actuates the bourdon tube. When mercury or its alloys are used as a filling medium, the temperature-volume-motion relationship is substantially linear. When hydrocarbon liquids are used, the liquid compressibility is appreciable, and the temperature-motion relationship is not so linear.

Since the fluids (liquid or gas) are homogeneous and extend to the bourdon tube, temperature changes on the capillary and on the bourdon tube will cause errors. These are made small by minimizing the volume in the capillary and bourdon tube and by providing ambient-temperature compensation. This compensation can be a duplicate system without a bulb to subtract the effect of the error; it can be a bimetallic compensator on the bourdon tube alone; or in the case of the mercury system a special capillary may be threaded with an invar wire compensator. The gas system has a relatively large bulb, a long range span (about 200°F minimum at room temperatures, 400°F near 1000°F), and the span may extend to a lower limit of about −400°F and an upper limit of about 1200°F. Hydrocarbon liquid systems have small bulbs, short range spans (as low as 25°F), and the span may extend to a lower limit of about −125°F and an upper limit of about 600°F. Mercury systems have somewhat larger bulbs (because of mer-

ature (in degrees Kelvin) of the solid at the peak of the curve, and T' is the temperature on the low-temperature side at which the emission is one-half its peak value. The figure shows *glow curves for* several zinc sulfide phosphors. The incorporation of *impurities* (activators, coactivators) into "host" compounds such as zinc sulfide is one of the most important processes for preparing solid *materials*; for a detailed discussion *see* LUMINESCENCE. When more than one type of trap is present, the glow curve consists of a corresponding number of peaks, which often may be resolved and analyzed as described. Thus thermoluminescence may be used to secure information about the properties of traps in solids. *See* TRAPS IN SOLIDS. [C.C.K.; J.H.S.]

Thermomagnetic effects

Electrical and thermal phenomena occurring when a conductor or semiconductor which is carrying a thermal current (that is, is in a temperature gradient) is placed in a magnetic field

Let the temperature gradient be transverse to the magnetic field H_z, for example along x. Then the following transverse-transverse effects are observed:

1. Ettingshausen-Nernst effect, an electric field along y,

$$E_y = Q(\partial T/\partial x)H_z$$

where Q is known as the Ettingshausen-Nernst coefficient. This coefficient is related to the Ettingshausen coefficient P by

$$P = QT\sigma$$

where σ is the thermal conductivity in a transverse magnetic field. This relation was discovered by P. W. Bridgman; it has been shown to be an example of the Onsager reciprocity relations of irreversible thermodynamics. *See* GALVANOMAGNETIC EFFECTS, THERMODYNAMICS (CHEMICAL); THERMOELECTRICITY.

2. Righi-Leduc effect, a temperature gradient along y,

$$(\partial T/\partial y) = S(\partial T/\partial x)H_z$$

where S is known as the Righi-Leduc coefficient

Also, the following transverse-longitudinal effects are observed:

3. An electric potential change along x, amounting to a change of thermoelectric power.

4. A temperature gradient change along x, amounting to a change of thermal resistance.

Let the temperature gradient be along H. Then changes in thermoelectric power and in thermal conductivity are observed in the direction of H

For related phenomena *see* HALL EFFECT; MAGNETORESISTANCE. [E.A.; F.K.]

Bibliography: See GALVANOMAGNETIC EFFECTS

Thermometer

An instrument that measures temperature. Although this broad definition includes all temperature-measuring devices, they are not all called thermometers. Other names have been generally adopted. For discussion of two such devices, *see* PYROMETER; THERMOCOUPLE. For general discussion of temperature measurement, *see* TEMPERATURE MEASUREMENT.

A variety of techniques are used in instruments known as thermometers. Some of these depend on the expansion of a liquid or metal for the indicating means. Others employ the change in pressure of a gas to detect the temperature. Still others use the change in electrical resistance which occurs with temperature changes.

Liquid-in-glass thermometer. This thermometer consists of a liquid-filled glass bulb and a connecting partially-filled capillary tube. When the temperature of the thermometer increases, the differential expansion between the glass and the liquid causes the liquid to rise in the capillary. In Fig. 1a the graduations are etched on the glass stem. The thermometer in Fig. 1b has a separate graduated scale similar to that of the common household thermometer. A variety of liquids, such as mercury, alcohol, and pentane, and a number of different glasses are used in thermometer construction, so that various designs cover diverse ranges between about −300°F and +1200°F. Expansion and contraction chambers are sometimes provided at each end of the capillary to permit over-range and un-

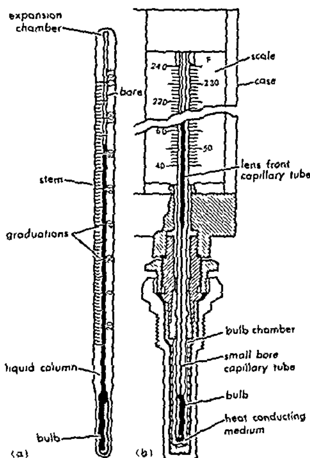


Fig. 1. Liquid-in-glass thermometers. (a) Etched-stem clinical thermometer (b) Graduated-scale industrial thermometer. (From D. M. Considine, ed., Process Instruments and Controls Handbook, McGraw-Hill, 1957)

metals, since they are stable, have a reason- temperature coefficient of resistance, and can awn into fine homogeneous wires with a high ance per unit length. Platinum can be used factorily between −258°C and 900°C, nickel een −150°C and 300°C, and copper between ?C and 120°C. Fine wires with a diameter bout 2.5 mils are used in making resistance ents, and these are wound on supporting tures of various shapes. In special cases, thin are used The bare-wire resistance thermom- responds rapidly to temperature changes, but se is restricted to clean, noncorrosive, noncon- ing gases. Usually, the resistance element must nserted in a protecting sheath or well, and its amic characteristics are not as good as those of ermocouple, which does not require electrical d hence thermal) insulation from the well A e constant in the neighborhood of 1 minute can expected from a protected resistance thermom r, but this varies widely depending upon the con- uction of the element and the characteristics of material being measured Industrial resistance rmometers usually have accuracies of 0.5°F thin the working range, but the error may be F and more at the range extremes. Carefully librated and maintained laboratory resistance rmometers may have an accuracy of 0.01°C.

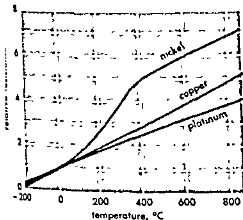


Fig 6 Typical relative-resistance curves of several metals used in resistance thermometers Relative resistance is the ratio of the resistance at the temperature of the metal to the resistance at °C (From D. M Considine, ed, Process Instruments and Controls Handbook, McGraw-Hill, 1957)

Since resistance thermometers carry a current, a self heating error occurs By keeping currents small and thermal conductivities high, this effect may be made negligible on most applications In dc thermometry, thermal emfs must be carefully avoided in the circuitry In ac thermometry, the circuitry must minimize inductive and capacitive disturbances

Thermistor. This device is made of a solid semiconductor with a high temperature coefficient of resistance. The thermistor has a high resistance

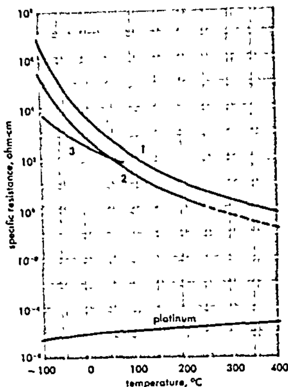


Fig 7 Resistance temperature characteristics of three typical thermometers. (From D. M. Considine, ed, *Process Instruments and Controls Handbook, McGraw-Hill, 1957*)

(see Fig 7), in comparison with metallic resistors, and is used as one element in a resistance bridge. *Since thermistors are more sensitive to temperature changes than metallic resistors,* accurate readings of small changes are possible. Thermistors are ceramic recrystalized mixtures of oxides of various metals and are usually in the form of small beads or disks with metallic leads. Thermistors are not as stable as metallic resistances, but certain compositions with good protection and care may change less than 1% per year In general, thermistors are used between 100 and 400°C. They drift or deteriorate at higher temperatures, and at low temperatures their resistance tends to become excessive See THERMISTOR. [R.E CL.]

Bibliography: D M. Considine (ed.), *Process Instruments and Controls Handbook*, 1957.

Thermonuclear reaction

A nuclear fusion reaction which occurs between various nuclei of the light elements when they are constituents of a gas at very high temperatures. Thermonuclear reactions, the source of energy generation in the sun and the stable stars, are utilized in the fusion bomb See FUSION, NUCLEAR; HYDROGEN BOMB, STELLAR EVOLUTION; SUN.

Thermonuclear reactions occur most readily between isotopes of hydrogen (deuterium and tritium), and less readily among a few other nuclei of higher atomic number At the temperatures and densities required to produce an appreciable rate

cury's low temperature coefficient of expansion) and longer range spans and are used at temperatures between −40°F and 1200°F. Normally, accuracies of 1% of the range span are obtained from these instruments, but this is achieved only by proper selection with full knowledge of application conditions.

Vapor-pressure thermal system. This filled-system thermometer utilizes the vapor pressure of certain stable liquids to measure temperature, as shown by Fig. 4. The useful portion of any liquid vapor pressure curve is between approximately 15

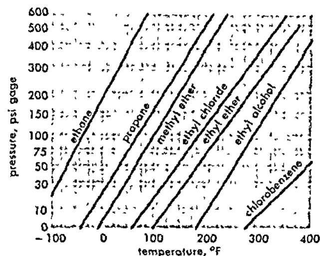


Fig. 4 Vapor-pressure vs temperature curves for various thermal system fills. (*D. M. Considine, ed., Process Instruments and Controls Handbook, McGraw-Hill, 1957*)

psi and either the critical pressure or the temperature at which the liquid begins to dissociate. A nonlinear relationship exists between the temperature and the vapor pressure, so the motion of the bourdon tube is greater at the upper end of the vapor-pressure curve Therefore, these thermal systems are normally used near the upper end of their range, and an accuracy of 1% or better can be expected. Vapor-pressure systems are designed so that the active liquid-vapor interface occurs in the bulb, and the effective temperature occurs at this interface. There is no error due to ambient temperature changes on the capillary, and only the temperature effect on the metal bourdon tube produces an error at this point. The bourdon-tube error is normally small and may be compensated (bimetallic) if it must be reduced.

When the bulb and the bourdon tube are not at the same level a hydrostatic error occurs, but this is easily removed by zero setting. The effect of atmospheric pressure variations is minimized by using only the elevated portion of the vapor-pressure curve of the various liquids. Range spans vary widely, but near room temperature the useful portion of the span is about 120°F, and at elevated temperatures it is 200°F. Few vapor-pressure systems are used below 0° and above 650°F.

The greatest advantage of the filled-system thermometer is its ability to provide a good, low-cost, temperature indication or record at a convenient point reasonably remote (up to 200 ft) from the temperature being measured. The bourdon tube is powerful enough to operate sensitive detectors, the output of which can be amplified pneumatically, electrically, or hydraulically for control purposes. The particular characteristics of each class of thermal system determine which will give the best service on various applications.

Resistance thermometer. In this type of thermometer the change in resistance of conductors or semiconductors with temperature change is used to measure temperature. Usually, the temperature-sensitive resistance element is incorporated in a bridge network which has a reasonably constant power supply. While a deflection circuit is occasionally used, almost all instruments of this class use a null-balance system, in which the resistance change is balanced and measured by adjusting at least one other resistance in the bridge. All of the resistors in the bridge, except the measuring resistance, have low temperature coefficients, and the entire bridge circuit is designed to be insensitive to ambient temperature effects. The power supply to the resistance thermometer may be either direct or alternating current, the former preferred for precision measurements and the latter preferred when a servo system is used to rebalance the bridge. Figure 5 shows an industrial resistance thermometer

Metals commonly used as the sensitive element in resistance thermometers are platinum, nickel, and copper, and the change in resistance per °C is illustrated in Fig. 6. These are the most satisfac-

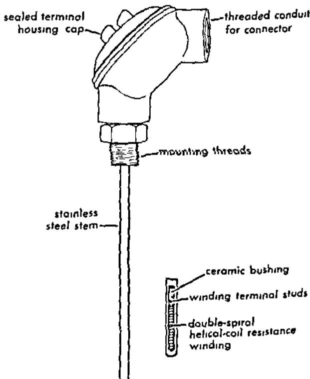


Fig. 5. Industrial-type resistance thermometer. (*From D. M Considine, ed., Process Instruments and Controls Handbook, McGraw-Hill, 1957*)

els, but lower limits vary widely. In small mammals with little protective insulation the lower limit may not exceed 0°C (32°F). Large animals with heavy fur generally are capable of withstanding the lowest temperatures on earth.

Homeothermy is partial as well as limited. The temperature of the central parts of the body (brain, heart, and abdominal viscera) are close enough to uniformity to justify the concept of a central core of uniform temperature, but the outer tissues, or shell, show a marked gradient in temperature. Thus, the temperature of the surface of the hand on a cold day may be 15°C (59°F), while the temperature on the skin of a seal in ice water may approach 0°C (32°F). The ability to maintain integrated function while parts of the body are at widely different temperatures constitutes one of the most remarkable and least understood properties of homeothermic animals.

Heat loss and heat production. As in all animals, the heat of the body is generated by chemical changes, notably oxidation, occurring in the tissues at large. The ultimate source of heat in the body is the oxidized food which liberates the same quantity of heat as it would if burned. The heat that is produced is lost through the feces and urine, by conduction and radiation from the skin, fur and feathers, and by evaporation of water from the skin and lungs.

Regulation of body temperature is accomplished by controlling both the rate at which heat is produced and the rate at which it is lost. When the external temperature is high and the temperature difference between core and ambient, or the excess temperature, is small, heat is lost in large measure by evaporation of moisture through panting (dog) or perspiration (man) and by a large flow of warm blood to the surface. The posture is adjusted to expose a maximal surface area.

When the external temperature is low and the temperature difference is large, heat loss by evaporation is greatly reduced, blood flow to the surface is restricted by closing of small blood vessels in the skin (vasoconstriction), the insulation is increased by erection of fur or feathers (piloerection), and surface area is minimized by assumption of a more spherical posture. This control of heat loss was termed physical regulation by M. Rubner.

If the ambient temperature is lower than that which can be compensated for by physical regulation alone, the animal increases heat production (cold thermogenesis). This is brought about by gross muscular exercise, by involuntary shivering, and by nonshivering mechanisms. Shivering is a fine muscle tremor in which most of the energy of contraction is converted into heat. Nonshivering heat production occurs without muscle contraction or tremor. Voluntary exercise is not as efficient as the involuntary mechanisms because it entails a greater energy cost owing to increased circulation, increased conduction from the surface, and greater surface exposure to cold.

The mechanisms controlling the loss and production of heat are largely dependent upon the external temperature. Within a range of temperature termed the zone of thermal neutrality, heat production can be maintained at a minimum, since variations in temperature are compensated for by regulation of heat loss. Below this zone, physical regulation is insufficient and heat production is increased, usually in proportion to the excess temperature between the deep body, or core, and the ambient. Ultimately, a limit to the heat production is reached at 3–5 times the thermoneutral rate and there is a failure in temperature regulation. Acclimatization of the individual to cold may extend this limit by increasing the capacity to produce heat, through enhancement of nonshivering heat production.

Species differences. There are very large species differences in both the range of thermal neutrality and in the minimum, or basal, metabolism. In lower animals and in those of small size there is only a very narrow range of environmental temperature over which metabolic rate is minimal. Mice, small birds, and shrews have a thermoneutral range between about 30°C (86°F) and 35°C (95°F), below which they must resort to increase in heat production. Muskrats, rabbits, cats, and dogs have a greater range of thermal neutrality extending down to 5°C (41°F) or lower, while arctic mammals of large body size do not increase their metabolism until the temperature is below −40°C. The critical temperatures at which metabolism must be increased in the cold are adaptive to climate. Tropical mammals generally have high critical temperatures, while arctic mammals generally have low critical temperatures. The latter are dependent on the insulation provided by thick fur which usually changes with the seasons. Arctic aquatic mammals also have low critical temperatures which are dependent on the insulation provided both by thick blubber and by special mechanisms of heat conservation in the appendages. In porpoises, parallel arrangement of arteries and veins in the appendages constitutes an effective heat exchanger system for returning the heat of the warm arterial blood to the body. Similar mechanisms are common in the limbs of nonaquatic birds and mammals, and help to extend the range of thermal neutrality within which basal metabolism can be maintained.

The basal metabolic rate, or BMR, of different species is closely related to body weight. Per unit weight, small animals have a much higher BMR than larger ones. Thus, an 18-gram (g) canary has about three times the BMR of a 300-g pigeon or rat, which in turn have about three times the BMR of a man. The accompanying correlation holds true when strict precautions of measurement are observed, such as thermal neutrality, avoidance of activity, and fasting. Under these conditions, the BMR of different species is proportional to the 0.7 power of the body weight or to the approximate surface area of the body. In man, the BMR is pro-

of thermonuclear reactions, all matter is completely ionized, that is, it exists only in the plasma state (*see* PLASMA PHYSICS). Thermonuclear fusion reactions may then occur within such an ionized gas when the agitation energy of the stripped nuclei is sufficient to overcome their mutual electrostatic repulsions, allowing the colliding nuclei to approach each other closely enough to react. For this reason, reactions tend to occur much more readily between energy-rich nuclei of low atomic number (small charge), and particularly between those nuclei of the hot gas which have the greatest relative kinetic energy. This latter fact leads to the result that, at the lower fringe of temperatures where thermonuclear reactions may take place, the rate of reactions varies exceedingly rapidly with temperature.

The reaction rate may be calculated as follows: Consider a hot gas composed of a mixture of two energy-rich nuclei, for example, tritons and deuterons. The rate of reactions will be proportional to the rate of mutual collisions between the nuclei. This will in turn be proportional to the product of their individual particle densities. It will also be proportional to their mutual reaction cross section σ and relative velocity v. Thus, the expression

$$R_{12} = n_1 n_2 (\sigma v)_{12} \text{ reactions}/(\text{cm}^3)(\text{sec})$$

gives the rate of reaction. The quantity $(\sigma v)_{12}$ indicates an average value of σ and v obtained by integration of these quantities over the velocity distribution of the nuclei (usually assumed to be maxwellian). Since the total density $n = n_1 + n_2$, then if the relative proportions of n_1 and n_2 are maintained, R_{12} varies as the square of the total nuclear particle density.

The thermonuclear energy release per unit volume is proportional to the reaction rate and the energy release per reaction

$$P_{12} = R_{12} W_{12} \text{ ergs}/(\text{cm}^3)(\text{sec})$$

If this energy release, on the average, exceeds the energy losses from the system, the reaction can become self-perpetuating. *See* CARBON NITROGEN CYCLE; KINETIC THEORY OF MATTER; MAGNETOHYDRODYNAMICS; NUCLEAR REACTION; PINCH EFFECT; PROTON-PROTON CHAIN [R.F.P.]

Bibliography: See FUSION, NUCLEAR.

Thermoregulation

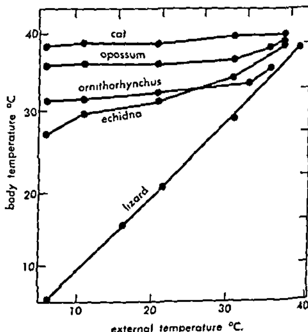
The property mammals and birds possess to maintain central body temperature at a constant level when exposed to variations in cooling power of the external medium. Animals that regulate body temperature are called homeotherms (constant temperature); the nonregulators are termed poikilotherms (varied temperature). Owing to maintenance of relatively high body temperatures in most homeotherms (37°C, or 98.6°F, in mammals and 41°C, or 106°F, in birds) and low body temperatures in most poikilotherms, the two groups have been loosely termed warm-blooded and cold-blooded animals, respectively. The terms, however, are inaccurate because it is the constancy rather than the levels of temperature that differentiates the two groups.

Homeothermy is an important example of the general emancipation of the higher organisms from the effects of changes in their environment. By maintaining a constant warm body temperature, homeotherms may be active at environmental temperatures which would produce torpidity in poikilotherms. This freedom, however, is gained only at the cost of maintaining a high rate of heat production, hence imposing high food requirements. The hibernators have partly solved this problem of energetics by periodically abandoning homeothermy during the coldest period of the year. Similarly, bats and hummingbirds abandon homeothermy when not active.

Scope and limits of temperature regulation. In insects, fishes, frogs, and lizards, body temperature varies almost directly with environmental temperature. Primitive regulation is seen in certain reptiles that orient their bodies to absorb maximum solar radiation when cool and seek shade when overheated; this is regulation by behavior. Automatic physiological temperature regulation is first seen in primitive mammals, like marsupials and monotremes, which tend to have low body temperatures and a narrow range of constancy (*see* *Ornithorhynchus* and *Echidna* in illustration).

In higher mammals, such as the cat, body temperatures are constant over a much wider range of ambient temperature, but there are limits of heat and cold in all animals, beyond which homeothermy cannot be maintained.

The upper limits of environmental temperature in most animals are fixed by body temperature lev-



Body temperature regulation. (From C. J. Martin, Thermal adjustments and respiratory exchange in monotremes and marsupials, Phil. Trans. Roy. Soc. London, Ser. B, 195, 1–37, 1902)

lower limits vary widely In small mam- ith little protective insulation the lower ay not exceed 0°C (32°F). Large animals avy fur generally are capable of withstand- lowest temperatures on earth.

eothermy is partial as well as limited. The ature of the central parts of the body , heart, and abdominal viscera) are close to uniformity to justify the concept of a core of uniform temperature, but the outer , or shell, show a marked gradient in tem- re. Thus, the temperature of the surface of and on a cold day may be 15°C (59°F), the temperature on the skin of a seal in ice may approach 0°C (32°F). The ability to ain integrated function while parts of the are at widely different temperatures consti- one of the most remarkable and least under- properties of homeothermic animals.

at loss and heat production. As in all ani- , the heat of the body is generated by chemi-

[illegible]

ntity of heat as it would if burned The heat is produced is lost through the feces and urine, conduction and radiation from the skin, fur , feathers, and by evaporation of water from the and lungs.

Regulation of body temperature is accomplished controlling both the rate at which heat is pro- ced and the rate at which it is lost When the ex- nal temperature is high and the temperature ference between core and ambient, or the excess mperature, is small, heat is lost in large measure evaporation of moisture through panting (dog) perspiration (man) and by a large flow of warm ood to the surface. The posture is adjusted to ex- se a maximal surface area.

When the external temperature is low and the mperature difference is large, heat loss by evapo- ation is greatly reduced, blood flow to the surface restricted by closing of small blood vessels in he skin (vasoconstriction), the insulation is in- reased by erection of fur or feathers (piloerec- ion), and surface area is minimized by assump- ion of a more spherical posture. This control of heat loss was termed physical regulation by M. Rubner

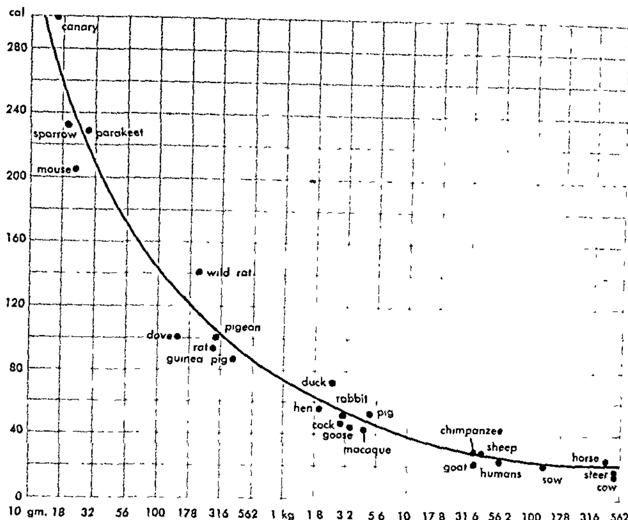
If the ambient temperature is lower than that which can be compensated for by physical regu- lation alone, the animal increases heat production (cold thermogenesis). This is brought about by gross muscular exercise, by involuntary shivering, and by nonshivering mechanisms. Shivering is a fine muscle tremor in which most of the energy of contraction is converted into heat. Nonshivering heat [illegible]

greater energy cost owing to increased circulation, increased conduction from the surface, and greater surface exposure to cold.

The mechanisms controlling the loss and produc- tion of heat are largely dependent upon the exter- nal temperature. Within a range of temperature termed the zone of thermal neutrality, heat produc- tion can be maintained at a minimum, since var- iations in temperature are compensated for by regulation of heat loss. Below this zone, physical regulation is insufficient and heat production is increased, usually in proportion to the excess temperature between the deep body, or core, and the ambient. Ultimately, a limit to the heat produc- tion is reached at 3–5 times the thermoneutral rate and there is a failure in temperature regulation. Acclimatization of the individual to cold may ex- tend this limit by increasing the capacity to produce heat, through enhancement of nonshivering heat production.

Species differences. There are very large spe- cies differences in both the range of thermal neu- trality and in the minimum, or basal, metabolism. In lower animals and in those of small size there is only a very narrow range of environmental tem- perature over which metabolic rate is minimal. Mice, small birds, and shrews have a thermoneu- tral range between about 30°C (86°F) and 35°C (95°F), below which they must resort to increase in heat production. Muskrats, rabbits, cats, and dogs have a greater range of thermal neutrality extending down to 5°C (41°F) or lower, while arctic mammals of large body size do not increase their metabolism until the temperature is below −40°C. The critical temperatures at which me- tabolism must be increased in the cold are adaptive to climate Tropical mammals generally have high critical temperatures, while arctic mammals gen- erally have low critical temperatures. The latter are dependent on the insulation provided by thick fur which usually changes with the seasons. Arctic aquatic mammals also have low critical tempera- tures which are dependent on the insulation pro- vided both by thick blubber and by special mecha- nisms of heat conservation in the appendages. In porpoises, parallel arrangement of arteries and veins in the appendages constitutes an effective heat exchanger system for returning the heat of the warm arterial blood to the body. Similar mechanisms are common in the limbs of non- aquatic birds and mammals, and help to extend the range of thermal neutrality within which basal metabolism can be maintained.

The basal metabolic rate, or BMR, of different species is closely related to body weight. Per unit weight, small animals have a much higher BMR than larger ones. Thus, an 18-gram (g) canary has about three times the BMR of a 300-g pigeon or rat, which in turn have about three times the BMR of a man. The accompanying correlation holds true when strict precautions of measurement are observed, such as thermal neutrality, avoidance of activity, and fasting. Under these conditions, the BMR of different species is proportional to the 0.7 power of the body weight or to the approximate surface area of the body. In man, the BMR is pro-



Basal metabolic rate, canary to cow (From F G. Benedict, Animal metabolism· from mouse to elephant, Sci in Progr, 1st ser., 1939)

portional to surface area. The latter can be calculated from the Dubois table of height and weight

Neuroendocrine control. The control of heat loss and heat production in response to change in external temperature is mediated through the nervous and endocrine systems There are receptors for heat and cold in the skin distributed over the entire surface of the body, which sense absolute temperature as well as rate and extent of change in temperature. Impulses from the receptors proceed to a central heat regulation center located in the hypothalamus, which is itself influenced by temperature. Sweating and panting can be induced directly by a rise in temperature of the center, while shivering can be elicited by a drop in temperature of the center and hence of deep body temperature.

Nervous impulses from the heat regulation center regulate the over-all response of the body to heat and cold as previously described. The concept of a center, with its supply of nerves from receptors, would lead one to assume that restriction of blood flow to the surface, or vasoconstriction, and increased heat production, or cold thermogenesis, are the result of the rate at which cold sensitive receptors are discharged, the number of receptors stimulated, and the thermal state of the center itself. Shivering is brought about directly through impulses from the center. In addition, impulses lead to increase in cold thermogenesis through the action of hormones particularly from the pituitary, adrenal, and thyroid glands *See* HIBERNATION; HOMEOSTASIS, HYPOTHERMIA; METABOLISM.

[J.S.H.]

Bibliography· F. G. Benedict, Animal metabolism: from mouse to elephant, *Sci in Progr.*, 1st ser., 1939, A C. Burton and O G. Edholm, *Man in a Cold Environment*, 1955; P F. Scholander et al., Heat regulation in some arctic and tropical mammals and birds, *Biol Bull*, 99·237–258, 1950

Thermosbaenacea

An order of small crustaceans in the superorder Pancarida. The order was erected in 1927 by T Monod and includes two families, the Thermosbaenidae and Monodellidae. Both families are monogeneric. *See* PANCARIDA [C B C.]

Thermostat

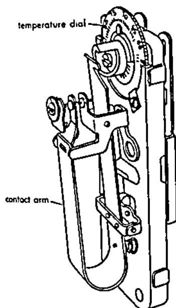
An instrument that responds to temperature, used as a temperature-controlling device The thermostat came into practical use in the mid-1880s. Early types operated the dampers of hand-fired heating furnaces and boilers in dwellings, and opened and closed individual steam radiator valves

. . . plants of larger buildings. Central

...

1917 the control industry was soundly estab-

In the decade following World War I, do- oil burners, stokers, gas burners, and elec- refrigerators revolutionized home heating and hold refrigeration; also unit heaters were oped for large space heating. Because these elements depended on temperature control for functioning, the demand for thermostats rapidly



Typical electrical contact-type thermostat (Minneapolis-Honeywell Regulator Co)

Successful air conditioning, following immediately, added tremendously to this demand and soon imposed a totally new requirement—the temperature regulation of individual operating mechanisms to ensure uniformity in the control of the conditioned space

...

tion, starts and stops oil burners, stokers, unit heaters, electrical heating elements, or refrigerators and moves dampers and automatic valves The second type, operating usually at a slower speed, can reverse direction before completing a cycle, or stop at a midpoint, then proceed in the same direction, or reverse. This type operates mixing valves and mixing dampers

Thermostatic control systems are pneumatic, electric electronic, or occasionally hydraulic

Thermostats in common use include the ordinary room type, both on and-off and positioning, the duplex, two separate units in one housing with manual or automatic selector switch for day or night operations at different control settings; summer-winter, with manual or automatic selector switch and set point change; immersion, for regulating liquid temperatures in tanks or piping; duct, for controlling flowing air temperature; limit, for safety stopping of the source when excessive temperatures threaten; and surface type for clamping on piping or securing to metal containers, to respond to and regulate the temperature of the contained fluid, also to prevent the operation of unit heaters when the steam supply in the piping fails.

Controlled items include the heat generators, refrigerators, blower and unit heater motors, automatic valves (on-and-off and mixing), automatic dampers (open-closed and mixing), pump, compressor and other equipment motors. *See* TEMPERATURE CONTROL, AUTOMATIC. [I.W C.]

Thermotherapy

The local application of heat to a part of the patient or generalized heating of the patient as a whole. If the objective of treatment of the whole patient is to induce an artificial elevation of bodily temperature, the treatment is referred to as hyperthermy Therapeutic use of abnormally low temperature is also included in this discussion.

Local application of heat. Local and systemic changes may result from local application of heat.

Local changes The first local effect is a local elevation of temperature. This reaches a maximum in about 20 min. Beyond this time the increased circulation of the blood in the region begins to carry away heat faster than it is being added (Fig 1).

The elevated temperature accelerates the chemical reactions of the tissue metabolism, resulting in more rapid formation of the metabolites which are the end products of these reactions. The engulfing of bacteria or debris by leukocytes in the region is also accelerated (*see* PHAGOCYTOSIS). Since the majority of the metabolites are acid, the acidity of the local tissues increases. Accompanying the rise

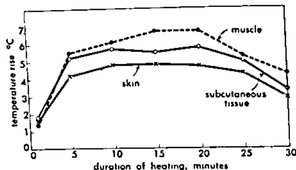


Fig 1 Effects on tissue temperature of exposure to microwaves (80 watts) The rise of temperature of the deeper tissues is greater than that of more superficial tissues. Tissue temperatures are lower after 30 min of heating than after 20 min of heating. (*From J. W. Gersten, K G. Wakim, J F Herrick, and F. H Krusen, Arch Phys Med., 30 7–25, 1949*)

of temperature, a reduction of the solubility of carbon dioxide (CO_2) in the tissue fluid causes an increase in the partial pressure of the carbon dioxide. The rise in CO_2 tension causes an increased dissociation of oxyhemoglobin and an increase in partial pressure of dissolved oxygen, thus liberating more oxygen. See HEMOGLOBIN.

Dilation of the arterioles occurs, with a secondary passive dilation of the capillaries that are supplied by the arterioles. The major cause of the dilation of the arterioles is probably the accumulation of metabolites, including carbon dioxide. It has been suggested, but not proved, that the heat may cause local formation of the histaminelike substance which acts to produce dilation of the arterioles by way of the axon or antidromic reflex. The dilation of the minute vessels leads to better nutrition in the region and better disposal of waste products. More leukocytes and antibodies are brought to the area. The hydrostatic pressure of the blood in the capillaries is increased, with a resultant increase in the transudation of fluid across the capillary wall into the tissue spaces. Accumulation of the fluid in the tissue spaces, if that occurs, constitutes edema. While thermotherapy does not ordinarily cause edema, it often aggravates preexisting edema or produces edema in a patient who is predisposed to it (see ANTIBODY).

Thermotherapy raises the threshold for painful stimulation. The mechanism of this analgesic effect may be related to a neutralization of the temperature gradient through the skin. With reduction of pain, muscles in the region may relax more. See PAIN, CUTANEOUS.

Systemic effects. Systemic or generalized effects of the local application of heat are related to the threat which it poses to the regulation of bodily temperature. In response to this threat, a dilation of cutaneous blood vessels occurs in areas remote from the part being treated. This increases the temperature of the skin in those areas and leads to greater loss of heat from the body (Fig 2). The stimulus for this reflex is the return of warmed blood from the area being heated, for if the venous flow of blood from the heated locality is obstructed, the reflex does not occur. The reflex center is the temperature-regulating center in the hypothalamus. The dilatation of many cutaneous vessels is a condition tending to produce a fall in the blood pressure, but that usually is prevented reflexly by the compensatory constriction of blood vessels in the gastrointestinal system and perhaps also in skeletal muscle and, if necessary, by an increased output of blood from the heart. Since the hands and feet are, respectively, the first and second parts of the body to participate in the regulation of bodily temperature by changes in cutaneous circulation, the reflex vasodilatation in response to remote heating occurs first in them. If the bodily region being heated is large enough to produce a significant increase in over-all metabolism, there will be compensatory increase in the ventilation of the lungs, represented by deeper and more rapid respiration. Another systemic effect of the local application of heat is sedation. The mechanism for this is unknown.

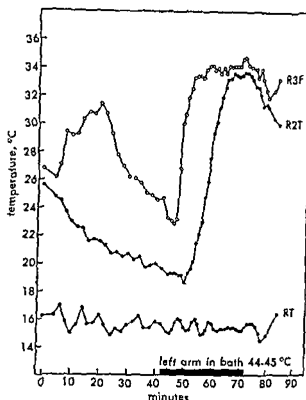


Fig. 2. Thermal reflex vasodilatation produced in the right hand and foot by immersing the left hand and forearm in hot water Vasoconstriction occurs earlier and to a greater extent in right foot than right hand prior to immersion of left forearm and hand in hot water and persists longer after the immersion (R3F = right third finger; R2T = right second toe; RT = room temperature) (From J. H. Gibbon, Jr and E. M. Landis, J. Clin Invest., 11(5).1019–1036, 1932)

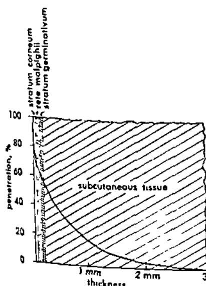
One of the effects of excessive dosage of heat locally applied is pain, which occurs when the temperature of the tissues reaches 45°C or more. With heat not quite so intense some of the arterioles in the area may become constricted, producing a mottled appearance known as erythema ab igne which may be followed ultimately by development of mottled pigmentation Since most of the beneficial effects of thermotherapy are associated with an increase in local circulation, it is better to keep the intensity of the heat below the level which will cause erythema ab igne. Heat which is much too intense may produce burns Patients having deficient sensation of pain or temperature are more likely to be burned than those whose protective mechanisms function normally Also, those having poor circulation to the skin are more likely to be burned than the normal person because they are unable to flush the area with large quantities of blood to convey the heat away.

Clinical methods of heat application. Any object in contact with the skin which has a temperature higher than that of the skin will transfer heat to the skin by conduction. Most commonly employed are the hot water bottle, hot packs, paraffin, and

cally heated pads and blankets. Hot wet usually are applied at temperatures from o 46.1°C. The Kenny packs, from which all moisture has been removed, may be applied at ratures in the neighborhood of 60°C because heat content, even at this temperature, is not ent to produce a burn. Paraffin is thinned mineral oil until it has a melting point of 57.2°C. *Because its heat content is only half* of water, it does not produce burns at this temperature. Chemical hot pads, which maintain their rature for about 1 hour, are rarely used now nductive heating devices heat only the areas which they are in contact, and are not very ble for irregular bodily surfaces. Furthermore, the exception of the chemical hot pad and ically heated pads, conductive heating de- undergo a relatively rapid drop in tempera- and because of this they heat inefficiently.

nvective heating is a special kind of conduc- heating. *The substance giving heat to the skin* ated *continually and is moved past the skin* nually, so that the temperature gradient be- n the heating substance and the skin is main- d more evenly. The whirlpool bath and related es are the only means for convective heating ommon use today The temperature of the water e whirlpool bath may be from 37.8 to 43.3°C. ddition to its heating effect, the whirlpool bath uently is *employed for its irrigating and gentle* saging action Use of hot air chambers at 1°C and devices for orificial heating, which ulate hot air or hot water at 51.7–54.4°C, is rare.

adiant heating devices employ the infrared e band. In this part of the spectrum, the skin is nearly a black-body radiator. The heat exchange is proportional to the difference in the fourth powers of the absolute temperatures of the heat emitter *and the skin. The quantum energy* in the infrared portion of the spectrum is not sufficient to induce more than an increase in the molecular and intramolecular motion which is heat. The near infrared portion of the electromagnetic spectrum extends from 770 to 1400 millimicrons (mμ), of which the 1200-mμ wavelength is the most penetrating. The far infrared portion of the spectrum extends from 1400 to 220,000 mμ (*see* ELECTROMAGNETIC RADIATION). The maximal penetration of infrared radiation into the skin is about 3 mm, but the far infrared portion of the spectrum penetrates only 0.05 mm (Fig. 3). Beyond 3000 mμ practically all of the energy is absorbed by *moisture on the skin and* does not reach the skin.



g. 3 Schematic diagram showing extinction of *infra-* d rays of the most penetrating wavelength (λ — 2 μ) by dead tissue The blood in living skin would use a much more rapid *extinction. (From J. D. Hardy* nd C. Muschenheim, *J Clin. Invest.*, 15(1) 1–9, 1936)

By means of conduction through the tissues and the convecting action of the circulating blood, the *heat is carried to deeper tissues than those penetrated by the infrared radiation. The intensity of* the heat reaching the skin is inversely proportional to the square of the distance from the source to the skin, and at a minimum when the rays are parallel to the skin (Lambert's cosine law).

In clinical practice, nonluminous infrared sources made of electrically heated wire or carborundum are employed to produce radiation in the *far infrared spectrum. Various* incandescent bulbs produce radiation in the near infrared spectrum and lesser amounts in the far infrared spectrum. Blackened glass bulbs, designed to cut down the glare of visible light, were discarded because of *their tendency to explode. Red glass bulbs* commercially supplied usually have too many optical imperfections in the lens, leading to the formation of "hot spots" that limit the general intensity of heat that may be applied to the skin. Clear glass or frosted bulbs are most satisfactory. These may be used as single bulbs with a built-in reflector or an external reflector, or they may be used in a heat *cradle* or "*baker*" *The baker provides an even* distribution of heat over a larger area than does a single bulb

Conversive heating devices may broadcase energy into the tissues via the radio or microwave frequency, or may transmit sound energy *in the ultrasonic* frequencies to the tissues via direct contact with the skin. The Federal Communications *Commission has set aside* frequency bands for the use of generators of short-wave diathermy and microwave diathermy. The use of generators of frequencies other than these is illegal, except in a completely shielded room.

The rapidly oscillating field of *diathermic currents* induces increased molecular and intramolecular motion. Aside from this production of heat, no *nonthermal or specific tissue effect* is known. Since some of the energy may not enter the patient, it is not possible to measure the dosage of diathermy and the only guide to dosage is the sensation of the patient For this reason its use is contraindicated

by a patient's deficiency of sensation Embedded metallic foreign bodies may produce arcing within the tissues or the concentration of the full electrical energy in their neighborhood, and are generally another contraindication to use of diathermy.

Ultrasound in frequencies of about 800,000 to 1,000,000 cycles per second is used clinically. Most of its effects are thermal, but there may be some effects due to cavitation produced by the rarefactive portion of the sound waves which are not yet clearly elucidated. Nerve fibers are damaged by ultrasonic energy more readily than are other tissues The sound energy is transformed partially to heat wherever it comes to an interface between two tissues having different acoustic impedances Reflection of the sound from bone tends to set up standing waves in the neighborhood of bone, with a greater concentration of energy in that region; and there is also some spread of energy along the bone from where the sound first strikes it. In its therapeutic applications ultrasound cannot be transmitted through air; from the soundhead of the generator it must be delivered to the skin by means of a coupling agent such as degassed water or mineral oil

Indications for thermotherapy. The physiologic effects of thermotherapy suggest the indications and contraindications for its use. Thermotherapy is used principally for its analgesic and sedative effects, and for increase of local circulation In the latter application it may accelerate the healing process and the resolution of infections. Thermotherapy is widely used as a treatment for various types of arthritis and painful conditions of muscles and ligaments, and as a preliminary adjunctive measure for exercise therapy

Hyperthermy. Hyperthermy, or treatment by artificial fever, is employed infrequently now, because most of the conditions for which it was used can be dealt with more efficiently and less expensively by antibiotics or by steroid therapy It still is valuable in some cases of rheumatoid arthritis and may be of value in treating some types of infection, such as generalized fungous infections.

Fever can be induced by diminution of heat loss from the body This can be accomplished by immersion in a hot bath (38 9–40.0°C) or by use of a fever therapy cabinet Deliberate infection with malaria to produce a fever gives rather unpredictable results, as does the injection of various foreign proteins such as typhoid vaccine. Hyperthermy must be administered by a highly trained therapist, familiar with the complications which may develop rapidly during the course of the fever.

The body attempts to combat the fever by generalized dilatation of the blood vessels of the skin. Because of this dilatation, the generalized increase of metabolism, and the direct effect of elevated temperature, there is an acceleration of heart rate and an increase in pulmonary ventilation. Profuse perspiration leads to loss of sodium chloride, which must be replaced by oral or intravenous administration of saline fluid. Suppression of perspiration can lead to heat stroke and possibly death. Heat exhaustion is a state similar to shock, more likely to occur than heat stroke.

Use of low temperatures. Refrigeration of a single extremity is employed sometimes when the blood supply to the extremity is inadequate to support the metabolism of the tissues at ordinary temperatures. The cooling may be achieved by packing the extremity in ice or wrapping it in a casing through which a refrigerant solution can be circulated. Low temperature has been used to inhibit the progression of gangrene in an extremity having inadequate circulation. Also, because low temperature blocks the passage of impulses through the peripheral nerves, it sometimes is employed as a method of anesthesia prior to the amputation of such a gangrenous limb. Generalized lowering of bodily temperature has been used extensively in recent years to reduce the metabolic demands of the tissues (particularly the brain, kidneys, and heart) so that the general circulation can be arrested safely for longer periods during operations on the heart In order to achieve this generalized chilling, it is necessary to administer enough sedative so that the shivering mechanism will be inhibited and will not interfere with the lowering of bodily temperature. The methods used for generalized cooling are similar to those used for cooling a single limb. After a surgical procedure the patient may be rewarmed by immersion in a warm water bath. *See* BIOPHYSICS [F.H.K ; G K.S]

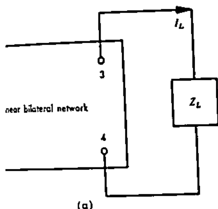
Bibliography: American Medical Association, *Handbook of Physical Medicine and Rehabilitation*, 1950; S. H. Licht (ed), *Therapeutic Heat*, 1958.

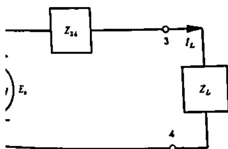
Thévenin's theorem (electric networks)

A valuable theorem in network problems which allows calculation of the performance of a device from its terminal properties only. It is not necessary to know the internal make-up of the device.

The theorem may be stated as follows: At any given frequency the current flowing in any impedance Z_L, connected to the terminals 3-4 of a linear bilateral network containing generators of the same frequency, is equal to the current flowing in the same impedance Z_L when it is connected to a voltage generator whose generated voltage is the voltage at terminals 3-4 with Z_L removed and whose series impedance is the impedance of the network looking back from terminals 3-4 into the network with all generators replaced by their internal impedances.

When a load impedance Z_L is connected across the output terminals 3-4 of the linear bilateral network of Fig. 1*a*, which contains generators of the same frequency, a current I_L will flow (Z_L may be considered as a branch of the network). Thévenin's theorem states that the circuit to the left of 3-4 in Fig. 1*a* may be replaced by a voltage generator (or a Thévenin generator) as in Fig. 1*b* where E_0 = open-circuit voltage determined at terminals 3-4 with Z_L removed, and Z_{34} = series impedance determined at terminals 3-4 with Z_L removed and





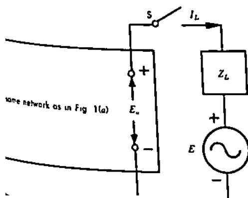
l Example of equivalent circuit for Thévenin's em (a) Actual circuit. (b) Equivalent Thévenin's t.

all generators replaced by their internal im- ınces. That is, the characteristics at the ter- als 3-4 are identical. The current in Z_L is

$$I_L = E_0/(Z_{34} + Z_L) \qquad (1)$$

: internal characteristics of the network, such power and efficiency, need not be identical.

he truth of Thévenin's theorem may be demon- ated in the following way. Imagine that a volt- : E is introduced in series with Z_L as in Fig. 2 is voltage E is of such magnitude and phase ation to E_0 (open-circuit voltage at 3-4) that = 0 when the switch S is closed, that is, E = . Then, by the superposition theorem, the zero



current may be said to consist of two equal and opposite currents, one of which is

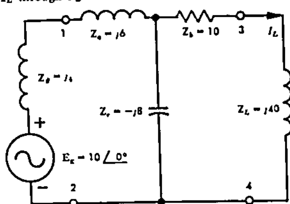
$$E/(Z_{34} + Z_L) = E_0/(Z_{34} + Z_L) \qquad (2)$$

Hence, the current that flows when only the actual generators are present is

$$I_L = E_0/(Z_{34} + Z_L) \qquad (3)$$

which is the same as Eq (1). *See* SUPERPOSITION THEOREM (ELECTRIC NETWORKS).

Thévenin's theorem is useful in complicated networks where Z_L is being varied, as in maximum power transfer. It should be remembered that the Thévenin circuit is equivalent only for the current I_L through Z_L.



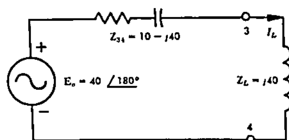


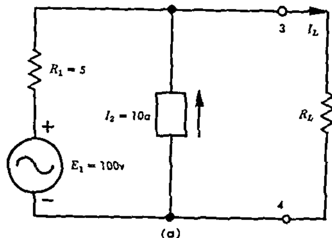
Fig 3 (a) Original network (b) Network with circuit to left of terminals 3-4 replaced by equivalent Thévenin's generator

The original network of Fig. 3*a* contains one generator having an $E_g = 10\angle 0°$ and an internal impedance $Z_g = 0 + j4$. Thévenin's theorem can be used to determine the equivalent voltage generator which can replace the actual generator and the actual network as far as the current I_L flowing in Z_L is concerned From this the current I_L can be calculated.

With Z_L removed, the open-circuit voltage E_0 at 3-4 is equal to the voltage across Z_c.

$$E_0 = \left[\frac{E_g}{Z_g + Z_a + Z_c}\right] Z_c = \frac{10\angle 0°}{j2}(-j8) = 40\angle 180°$$

The impedance determined at terminals 3-4 with



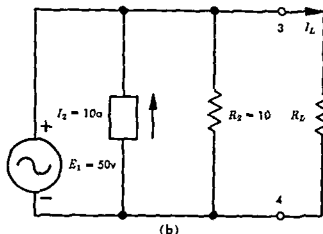


Fig. 4.

Z_L removed and the generator replaced by its internal impedance Z_g is

$$Z_{34} = Z_b + \frac{(Z_a + Z_g)Z_c}{Z_a + Z_g + Z_c} = 10 + \frac{(j10)(-j8)}{j2}$$
$$= 10 - j10$$

The Thévenin generator is shown to the left of 3-4 in Fig. 3b

The current is

$$I_L = \frac{E_0}{Z_{34} + Z_L} = \frac{40/180^\circ}{10 + j0} = 4/180^\circ$$

Figure 4 shows two dc circuits, each containing a voltage generator E_1 and a current generator I_2. As far as I_L is concerned, the portion to the left of 3-4 can be replaced by a Thévenin generator in each circuit.

For the circuit in Fig. 4a, with R_L removed, the open-circuit voltage E_0 at 3-4 is

$$E_0 = E_1 + R_1 I_2 = 100 + 50 = 150 \text{ volts}$$

The impedance Z_{34} determined at 3-4 with R_L removed and the generators replaced by their internal impedances, $R_1 = 5$ for the voltage generator E_1 and $R_2 = \infty$ for the current generator I_2, is

$$Z_{34} = R_1 = 5$$

In Fig. 4b with R_L removed, the open-circuit voltage E_0 at 3-4 is

$$E_0 = E_1 = 50 \text{ volts}$$

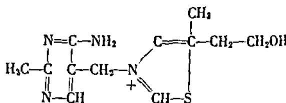
The impedance Z_{34} determined at 3-4 with R_L removed and the generators replaced by their internal impedances, $R_1 = 0$ for E_1 and $R_2 = 10$ for I_2, is

$$Z_{34} = 0$$

The current in each case would be $E_0/R_L + Z_{34}$ [K.Y.T.]

Thiamine

A water-soluble vitamin found in many foods; pork, liver, and whole grains are particularly rich sources. It is also known as vitamin B_1 or aneurin. Its structural formula is



The vitamin is heat labile, and considerable amounts are destroyed during cooking. Thiamine is unstable in alkaline solutions but stable in acid solutions It acts like a weak base and can be absorbed on basic ion-exchange materials such as decalso and fuller's earth, a property used in its chemical determination. Biological and microbiological methods for its estimation are available but are seldom used, and a chemical assay based on the production of thiochrome, a fluorescent reaction product of thiamine, is the method of choice. A dietary source of thiamine is required by all animals that have been studied. Thiamine deficiency is known as beriberi in humans and polyneuritis in birds.

Thiamine functions in enzyme systems as thiamine pyrophosphate, a coenzyme known as cocarboxylase Thiamine containing enzymes decarboxylate α-keto acids such as pyruvic acid and α-ketoglutaric acid. Thiamine pyrophosphate is also involved in the transketolation of pentose to heptulose in the hexose monophosphate shunt system which is an alternative to the conventional glycolysis pathway of anaerobic glucose metabolism. *See* CARBOHYDRATE METABOLISM.

Muscle and nerve tissues are affected by the deficiency, and poor growth is observed. People with beriberi are irritable, depressed, and weak. They often die of cardiac failure Wernicke's disease observed in alcoholics is associated with a thiamine deficiency This disease is characterized by brain lesions, liver disease, and partial paralysis particularly of the motor nerves of the eye As is the case in all B vitamin diseases, thiamine deficiency is

usually accompanied by deficiencies of other vitamins.

Thiamine is the most poorly stored of the B vitamins. Individuals eating vitamin-deficient diets are likely to develop beriberi symptoms first. Approximately 5 mg of thiamine can be absorbed per day by normal adults. Excess thiamine given by mouth or parenterally is usually lost through excretion in the urine and feces. Thiamine requirements are related to caloric intake. More thiamine is required when high carbohydrate diets are fed than when high fat diets are eaten, but the reason for this is still not clear. Some foods, particularly raw fish, contain enzymes which destroy thiamine. More thiamine is needed in such alterations in physical state as hyperthyroidism, pregnancy, and lactation. Thiamine requirements of humans are primarily estimated by means of urinary excretion data. The recommended dietary allowances of the National Research Council provide 0.5 mg of thiamine for each 1000 calories for adults, with a minimum of 1 mg per day. [S N C]

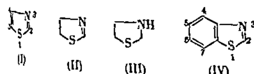
MANUFACTURE OF VITAMIN B_1

Industrial synthesis of this vitamin is accomplished by linking chloromethylpyrimidine with 4-methyl-5-(β-hydroxyethyl)-thiazole to give aneurin. Another way to build up the thiamine molecule

ring via a methylene bridge (U.S. Patents 2,193,858 and 2,218,350.) See VITAMIN. [F.D M.]

Thiazole

One of a class of organic heterocyclic compounds (sometimes specified as 1,3-thiazoles) in which a five-membered diunsaturated ring contains one atom of nitrogen and, in a nonadjacent position, one atom of sulfur. See AZOLE; HETEROCYCLIC COMPOUNDS. The preferred numbering is shown in formulation (I); however, since other numbering



systems have been employed, caution should be exercised in translating names to structures. Dihydrothiazoles are called thiazolines, of which the most familiar are the Δ^2-thiazolines (II). Tetrahydrothiazoles are called thiazolidines (III). Thiazole fused to a benzene ring is benzothiazole (IV). Important thiazole derivatives include vitamin B_1 (thiamine) and sulfathiazole. Several valuable dyes and rubber-vulcanizing accelerators contain the benzothiazole nucleus. Penicillin is a thiazolidine derivative.

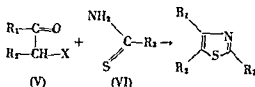
Properties. The thiazole ring is a resonance-stabilized aromatic system. The ring is relatively resistant to hydrolysis and to disruptive oxidation with nitric acid. Bromine-water as well as permanganate, however, do oxidize the ring. Thiazole can undergo electrophilic substitution such as nitration and sulfonation at the 4 and the 5 positions. Further, in line with the aromatic character of thiazole, aminothiazoles form diazonium salts with nitrous acid. See AROMATIC HYDROCARBON.

The parent compound, thiazole (I), bp 117°C, is a colorless, water-soluble liquid with an odor resembling that of pyridine. Although thiazole is a weak base (pK_a 2.53), acids form simple salts, and alkylating agents form quaternary salts. Quaternary thiazolium salts, although stable in acid, are decomposed in alkali and ring rupture occurs.

Amino groups at position 2 can be diazotized. Subsequent coupling to give azo compounds, and replacement of the diazonium group by hydrogen, halogen, hydroxyl, or nitro in standard procedures are possible. The SH or mercapto group at position 2 can be replaced with hydrogen either by treatment with acidic hydrogen peroxide, or by desulfurization with Raney nickel.

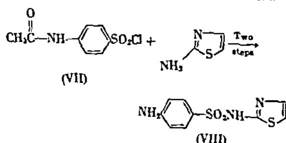
The chemistry of the thiazole 2 position—but in contrast not that of the 4 position—is similar to the chemistry of the pyridine 2 position. Thus, the 2-halothiazoles are convertible with relative ease to 2-hydroxy, 2-alkoxy, 2-amino, 2-mercapto, and by reduction, to 2-hydro derivatives. Thiazole-2-carboxylic acid decarboxylates with relative ease. Further, 2-methylthiazole metalates at the methyl group without difficulty, and also condenses with benzaldehyde.

Preparation. The most versatile general synthesis of thiazoles condenses α-halo carbonyl compounds (V) with thioamides (VI). By suitable



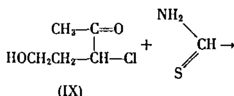
choice of the R groups, many kinds of thiazoles can be prepared. The thioamides (VI) used most frequently are thioformamide (R_3 = H), thiourea (R_3 = NH_2), and dithiocarbamate ion (R_3 = SH).

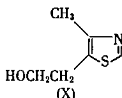
Sulfathiazole (VIII) is one of the useful bacteriostatic sulfa drugs. Reaction of p-acetamidobenzenesulfonyl chloride (VII) with 2-aminothiazole



zole gives the *N*-acetylated sulfathiazole, which on alkaline hydrolysis forms sulfathiazole.

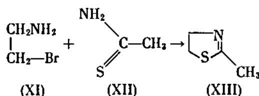
Thiamine includes as part of its structure, the thiazole derivative (X). A practical synthesis of this fragment proceeds by condensing the α-chlorinated derivative (IX) of 5-hydroxy-2-pentanone



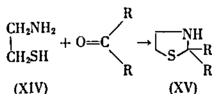


with thioformamide. The thiazole (X), combined with the proper pyrimidine moiety, gives thiamine.

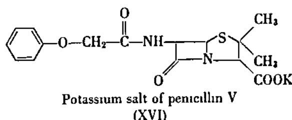
Δ^2-Thiazolines, for example 2-methyl-Δ^2-thiazoline (XIII), are prepared by combining a β-bromoamine (XI) with a thioamide (XII). Thiazolidines (XV) form when aldehydes or ketones react



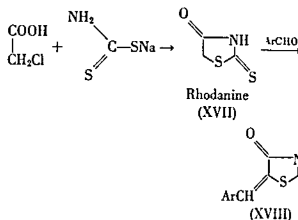
with β-amino mercaptans (XIV). In the laboratory



synthesis of penicillin, the thiazolidine ring (XVI) is formed by this process. Rhodanine, or 4-ketothiazolidin-2-thione (XVII), is prepared by the reaction of chloroacetic acid with dithiocarbamate. Condensation of rhodanine with carbonyl compounds give 5-methylene derivatives (XVIII),

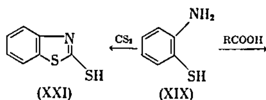


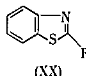
Potassium salt of penicillin V
(XVI)



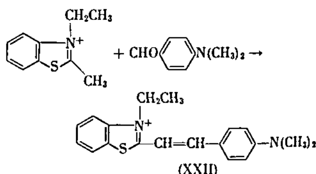
which can be converted to a variety of useful products.

Benzothiazole syntheses start with *o*-aminothiophenol (XIX), which, with carboxylic acids, gives

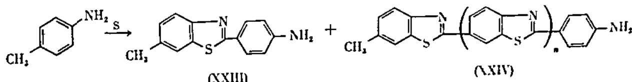




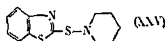
2-substituted benzothiazoles (XX), and with carbon disulfide, gives 2-mercaptobenzothiazole (XXI). Cyanine dyes, for example (XXII), are



formed from the reaction of an aldehyde such as *p*-dimethylaminobenzaldehyde with quaternized 2-methylbenzothiazole Other benzothiazole dye materials are obtained when *p*-toluidine is fused with sulfur A mixture of dehydrothiotoluidine (XXIII) and "Primuline dye bases" (XXIV) is formed. Sulfonated "Primuline dye bases" (XXIV) is a commercial yellow dye which is used after diazotization by application to cotton. Other types of dyes have also been derived from benzothiazoles (XXIII) and (XXIV)



A commercial preparation of Captax or 2-mercaptobenzothiazole (XXI), a material of considerable value as an accelerator in the vulcanization of rubber, proceeds by heating aniline, carbon disulfide, and sulfur to 200–300°. Benzothiazole-2-sulfenamides (XXV), which are derived from 2-



mercaptobenzothiazole, are also effective accelerators. *See* DYE, ORGANOSULFUR COMPOUND; SULFA DRUGS. [W.J.GE.]

Bibliography: R. Adams (ed.), *Organic Reactions*, vol. 6, 1951, H. T. Clarke, J. R. Johnson, and R. Robinson (eds.), *The Chemistry of Penicillin*, 1949, R. C. Elderfield, *Heterocyclic Compounds*, vol. 5, 1957.

Thickening

The production of a concentrated slurry from a dilute suspension of solid particles in a liquid. In practice, a thickener is usually expected also to produce a clear liquid, and therefore thickening includes clarification as a concurrent objective. Thickening and clarification are applications of sedimentation, and both are representative of a larger group of industrial processes called mechanical separations.

Thickening may be accomplished either in batch equipment or in continuous units. The latter is more common. In continuous equipment, special means are needed to move the concentrated slurry to the outlet, and plants for this purpose are called

shaft. The tank may have either a flat or shallow cone bottom. Mechanically agitated thickeners may be quite large. Their major dimensions are 20–300 ft in diameter and 8–12 ft in depth. Small tanks

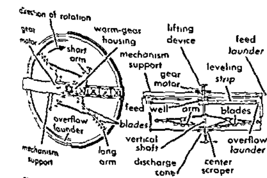


Fig. 1. Dorr thickener. (*From J. H. Perry, ed., Chemical Engineers' Handbook, ... :Graw-Hill, 1950*)

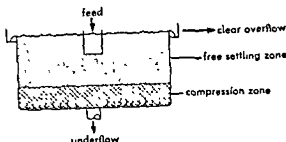


Fig. 2. Zones in continuous thickener. (*From W. L. McCabe and J. C. Smith, Unit Operations of Chemical Engineering, McGraw-Hill, 1956*)

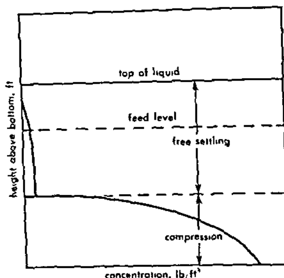


Fig. 3. Solid concentrations in continuous thickener. (*From W. L. McCabe and J. C. Smith, Unit Operations of Chemical Engineering, McGraw-Hill, 1956*)

Thickeners are especially useful when large volumes of dilute slurries must be treated, as in the manufacture of cement, the production of magnesium from sea water, the treatment of sewage, and the purification of water.

In operation, dilute feed pulp is admitted continuously through a launder to a central well immersed to a depth of 2–3 ft below the surface of the liquid in the tank. The liquid from the feed moves radially to the wall of the tank and overflows across the edge of the tank into a trough or launder which circles the periphery of the tank. During the flow of the liquid, the solids brought in with it settle, so that by the time the liquid reaches the overflow launder, it is free from solid. The solids settle to the bottom of the tank and form a concentrated slurry. The rakes, without repulping the solids into the liquid, gently agitate the solids, break up the flocs to aid the process of concentration, and move the thickened solids to the discharge in the center of the tank bottom. From the discharge, the thickened slurry flows to the suction of a sludge pump.

Figure 2 shows the zones in a continuous thickener, and Fig. 3 shows the relation between the concentration of solids and the height above the

bottom of the tank. Two main zones exist which are separated by an interface at which the rate of change, with vertical distance, of the concentration of solids increases sharply in the direction top to bottom. The upper zone, in which clarification is accomplished, is free from solid in its top layers and supplies the clarified liquid overflow. Within the clarification zone, the solid concentration varies from zero to a small value at the interface between the zones. In the clarification zone, the solid particles are sufficiently far apart that free settling takes place. The bottom zone, in which thickening is accomplished, contains most of the inventory of solid in the tank. The concentration of solid changes rapidly from that in the lowest level in the clarification zone to that of the thickened slurry leaving the thickener. The process occurring in this zone is essentially that of compression. See SEDIMENTATION (INDUSTRIAL).

To obtain satisfactory capacities, the feed to a thickener is often flocculated. The performance of a given plant operating on a given feed slurry depends largely on the major dimensions of the tank. To obtain a clear overflow, the upward velocity of the liquid in the clarification zone must be less than the minimum terminal settling velocity of the smallest particles. Since the velocity of the liquid is proportional to the horizontal cross section of the tank, the clarification capacity is approximately proportional to this area, and therefore, to the square of the diameter of the tank. The solid concentration in the underflow, and hence the degree of thickening achieved, depends on the time allowed for action in the compression zone. Once the feed rate of dilute slurry is fixed, the time for compression is proportional to the height of the compression zone. The performance of the unit as a thickener is, then, a function of the depth of the tank.

Larger capacities for a given floor area are obtainable in multitray thickeners. A single vertical steel tank is subdivided into a stack of individual settling compartments by shallow conical trays. Each tray is fitted with a rake-agitator which moves the sludge to the center draw-off of the tray. The thickened solids move down from tray to tray, and the final sludge, at maximum concentration, is discharged from the center of the bottom tray. The clear liquids overflowing from all trays are combined into a single overflow stream. Since clarification capacity is proportional to the total cross-sectional area of the settling zone, the ratio of the capacity of a multitray thickener to that of a single-stage unit of the same floor area is approximately equal to the number of stages. See CLARIFICATION; SEPARATION (MECHANICAL). [W.L.M.]

Bibliography: W. L. Badger and J. T. Banchero, *Introduction to Chemical Engineering*, 1955; E. W. Comings, Thickening calcium carbonate slurries, *Ind. Eng. Chem.*, 32:663, 1940; J. H. Perry (ed.), *Chemical Engineers' Handbook*, 3d ed., 1950; A. F. Taggart, *Handbook of Mineral Dressing*, 1945.

Thigh

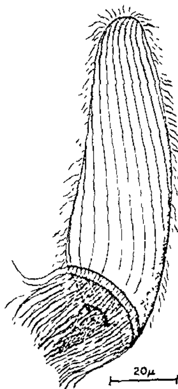
In vertebrates, the region between the hip and the knee. The bony support of the thigh is the large, weight-bearing femur, or thigh bone, which is surrounded by four muscular compartments in man. These contain functional groups for movements involving the body, thigh, and leg. Each is innervated by specific nerves arising from the lumbodorsal plexus. See LEG.

Connective tissue surrounds other structures and also forms the ligamentous attachments of the thigh to the hip and knee joints. The arteries arise principally from the femoral artery. Blood returns to the body through both a superficial and a deep set of veins which drain primarily into the femoral vein. Subcutaneous tissue and skin cover the thigh and are traversed by superficial nerves and vessels.

Analogous structures are present in most vertebrates, but less specialization is seen in lower tetrapods. [E.G.ST.]

Thigmotrichida

A restricted group of forms comprising an order of the Holotricha and generally found in association with mollusks from both fresh and salt water. Some are mouthless. Those species with a cytostome are generally equipped with a buccal ciliature which indicates an advance over the primitive hymenostome arrangement. This structure is located at, or near, the posterior pole of the organism's body. From an evolutionary point of view, it is postulated that thigmotrichs are the connecting link between certain hymenostomes and the peritrichs. *Boveria* (see illustration) and *Hemispeira*



Boveria, an example of a thigmotrichid.

are common genera. *See* HOLOTRICHA; *see also* HYMENOSTOMATIDA; PERITRICHIDA. [J.O C.]

Thinner

are expected to evaporate after application. Because their only function is to make the application simple, it is important that their cost be low. Water is used as a thinner in emulsion paints and in certain water soluble paints such as water colors and calcimines.

Petroleum fractions are most commonly used for oil and resin coatings. The fraction boiling between 300 and 400°F, called mineral spirits, is most widely used. A lower-boiling, and faster-evaporating, solvent is called VM&P (Varnish Makers' and Painters') naphtha. Still faster-evaporating materials are called petroleum ether, lacquer diluent, or rubber solvent. Stronger solvents contain substantial amounts of aromatic hydrocarbons and may be derived from petroleum or coal tar. These may be essentially pure materials, such as toluene or xylene, or mixtures designed to have the solvency and evaporation characteristics desired. *See* SOLVENT

Since numerous coatings resins are not sufficiently soluble in hydrocarbons, other materials or mixtures must be used. These include alcohols such

perature of the surface, and under humid conditions, they may cause moisture to condense on the surface, producing the phenomenon known as blushing

Historically, the thinner used for conventional paints was turpentine, but because of newer and cheaper solvents, it has largely disappeared from paint manufacturing, although it is still used to some extent for thinning paints on the job. *See* SURFACE COATING; TURPENTINE. [F SO]

Thio compounds

Organosulfur compounds in which one or more sulfur atoms replace oxygen. Because of the frequent use of this nomenclature and the large number of structural variations involved, the thio names may appear confusing but are generally understandable in specific cases. Thiols and thiophenols are the sulfur analogs of alcohols and phenols. Di- and trithio compounds have 2 or 3 sulfur atoms, corresponding to the oxygen analogs. Acyl SH compounds are the thio acids, for example, thioacetic acid, $CH_3C(=O)SH$, of which many are known, and which can be made (1) by acylating H_2S, or its salts, (2) via $RC(=O)OH$ and P_2S_5, or (3) from RMgX and COS (carbon oxysulfide). Dithio acids, $R-C(=S)-SH$, may, for example, be made from RMgX and carbon disulfide, CS_2 (analogous to the reaction of Grignard reagents with CO_2). Trithiocarbonates are derivatives of $S=C(SH)_2$, the sulfur analog of carbonic acid. Dithiocarbamates, related to carbamic acids, are readily obtained in the form of salts (the free acids are unstable) by the reaction, $2R_2NH + CS_2 \rightarrow R_2NC(=S)S^-(H_2N^+R_2)$. Carbon disulfide is important technically for such reactions and also for reactions with alcohols and alkali to give xanthates



$$ROH + CS_2 + KOH \rightarrow RO-C(=S)S^-K^+ + H_2O$$

for example, ethanol yields potassium ethyl xanthate. The synthesis of thiocarbonates often starts from phosgene, $O=CCl_2$, in which the reactive chlorine atoms can be replaced with sulfur groups, by reactions with RSH, RC(O)SH, and HSH

Many other compounds which carry the thio name are oxygen analogs. For example, thiocyanates, $RSC\equiv N$; isothiocyanates, $R-N=C=S$; thioureas, $(NH_2)_2C=S$; thioamides, $RC(=S)NH_2$; dithioesters, $RC(=S)SR'$; thio acid chlorides,



$$RC(=S)Cl$$

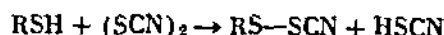
Oxidation of the thiocarbamates yields corresponding disulfides, known as thiuram disulfides,



$$R_2N-C(=S)-S-S-C(=S)NR_2$$

some of which are highly effective vulcanization accelerators

Thiocyanogen $(SCN)_2$ and chlorothiocyanogen $(Cl-SCN)$ are also of interest. The former has properties similar to the halogens and hence is useful as a thiocyanating agent, as in the preparation of sulfenyl thiocyanates:



$$RSH + (SCN)_2 \rightarrow RS-SCN + HSCN$$

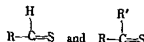
Thiobarbiturates are barbituric acid derivatives, in which, generally, an RS group has replaced an alkyl or aryl radical in the 2 position of barbituric acid

The thio acids, dithio acids, thiocarbamates, dithiocarbamates, thiocarbonates, trithiocarbonates, polythiols, thiocyanates, and many related compounds have a broad and useful set of chemical properties which have been reasonably well developed. Many of these compounds have already found extensive industrial applications *See* MERCAPTAN; ORGANOSULFUR COMPOUND; RUBBER.

[N.K.]

Thioaldehyde and thioketone

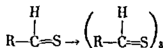
Organosulfur compounds of structure



$$R-\overset{H}{\overset{|}{C}}=S \quad \text{and} \quad R-\overset{R'}{\overset{|}{C}}=S$$

They are isolable only in rare instances because,

unlike the oxygen analogs (aldehydes and ketones), thioaldehydes and thioketones tend strongly to polymerize, for example,



The trimer is a trithiane derivative, with alternate C and S atoms in a 6-membered ring, each carbon of which bears R and H groups. The monomeric thiocarbonyl compounds are highly colored. The thiocarbonyl (—C═S) group also is found in thiourea and in heterocyclic thiones, but here tautomeric structures in which —C═S becomes ═C—SH are generally involved Hence, they are not true thiocarbonyl compounds. *See* ALDEHYDE; KETONE; ORGANOSULFUR COMPOUND. [N.K.]

Thiobacteriaceae

A family of nonfilamentous, gram-negative bacteria, of the suborder Pseudomonadineae, able to oxidize inorganic sulfur compounds. Although widely distributed in nature, Thiobacteriaceae are concentrated where hydrogen sulfide occurs Six genera are grouped in the Thiobacteriaceae, not because they are closely related, but for convenience The genera are *Thiobacterium, Thiogloea, Macromonas, Thiovulum, Thiospira*, and *Thiobacillus*. Members of the first five have not been grown in pure culture, and their physiology is virtually unknown. Hydrogen sulfide, free sulfur, and inorganic sulfur compounds, like thiosulfates, are oxidized to sulfuric acid by Thiobacteriaceae. This oxidation probably represents their respiration and provides energy for the fixation of carbon dioxide, although such carbon dioxide fixation has been demonstrated only in *Thiobacillus*. *See* BACTERIAL METABOLISM

Thiobacterium. These are rod-shaped nonmotile bacteria, 0.3–1.5 × 1.0–5.0 μ, forming small floating colonies on the surface of sulfide-containing waters. The colonies appear white by reflected light as the result of sulfur deposition inside or outside the cells. Individual cells may be embedded in a pellicle or in zoogloeal masses.

Thiogloea. These are nonmotile bacteria, varying in shape from spherical to ellipsoid and in size from 0.6 × 2–6 × 10 μ. The bacteria are at times entirely filled with droplets of amorphous sulfur. The droplets may be as large as 4 μ in diameter in the largest species. These bacteria form colonial aggregates in zoogloeal masses containing from less than 100 to several thousand individuals.

Macromonas. These cylindrical or bean-shaped motile bacteria, 3–14 × 8–30 μ, have a single polar flagellum 10–40 μ long One to four spherical inclusions of calcium carbonate may nearly fill the cell, small sulfur globules may also be found inside the cell.

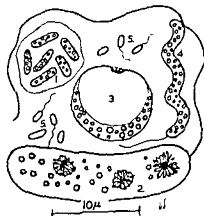
Thiovulum. These nearly spherical organisms are 5–20 μ in diameter. The cytoplasm is commonly concentrated near one end of the cell and may contain sulfur globules. The remainder of the cell is occupied by a vacuole. Rapid tumbling motion suggests polar flagellation, but no flagellum has been demonstrated.

Thiospira. This group includes slightly bent or twisted motile rods, about 2 μ wide in the center and 7–50 μ long, with one or two flagella at each end. The cells contain globules of sulfur and some metachromatic material.

Thiobacillus. These rod-shaped bacteria are either motile, with a single polar flagellum, or nonmotile. The bacteria are 0.5 × 1–3 μ in size. Members of this genus have been studied in pure culture. They are autotrophic organisms, using carbon dioxide as the sole source of carbon and deriving their respiratory energy from the oxidation of sulfide, sulfur, thiosulfate, polythionates, and, in the case of *T. thiocyanoxidans*, thiocyanate. One species, *T. novellus*, can live either by sulfur oxidation and CO_2 fixation or on organic substrates, as a facultative autotroph, while the other species are obligatory autotrophs being unable to utilize organic matter The oxidation of sulfur compounds requires oxygen, except that one species, *T denitrificans*, can oxidize sulfur compounds anaerobically if nitrate is present, the latter being reduced to nitrogen The fixation of carbon dioxide proceeds through reactions similar to those which participate in photosynthetic carbon dioxide fixation (*see* PHOTOSYNTHESIS).

The genus also includes *T. thioparus*, the type species, distinguished by its deposition of molecular sulfur when grown on a medium containing more than 0.5% sodium thiosulfate, and *T. thiooxidans*, which is favored by a medium of pH 2–3 5, while all other species prefer pH 7–8. Morphologically, all species are indistinguishable, except that *T. novellus* is frequently wider (0 4–0 8 μ) than other species (0 5 μ). Additional species have been reported. *See* BACTERIA, TAXONOMY OF, PSEUDOMONADINEAE, SCHIZOMYCETES [W.V.]

Bibliography: Z. Devidé, Zwei neue farblose Schwefelbakterien, *Schweiz. Z Hydrol.*, 14:446–455, 1952; W Vishniac and M. Santer, The Thiobacilli, *Bacteriol Rev.*, 21(3):195–213, 1957.



1–Thiobacterium 2– Macromonas
3– Thiovulum 4– Thiospira
5– Thiobacillus

Some genera of the Thiobacteriaceae. (V. B. D. Skerman)

Thiocyanate

One of a group of compounds, both organic and inorganic, which contain the —SCN group and are derived from thiocyanic acid, HSCN. Like cyanic acid, thiocyanic acid may exist in two forms, H—S—C≡N and S=C=N—H. The latter is called isothiocyanic acid and gives rise to isothiocyanates, which are well characterized as a class of organic compounds.

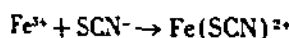
... and copper salts), and many complexes are formed with excess thiocyanate; for example, $[Pt(SCN)_4]^{2-}$ and $[Pt(SCN)_6]^{2-}$.

Potassium thiocyanate can be used to titrate Ag^+ as in the Volhard titration:



$$Ag^+ + SCN^- \rightarrow AgSCN \quad \text{(white precipitate)}$$

An excess of reagent is detected by the formation of a red complex of iron:



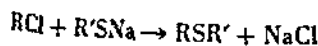
$$Fe^{3+} + SCN^- \rightarrow Fe(SCN)^{2+}$$

This latter reaction is used as a very sensitive test for both CNS^- and Fe^{3+} ions.

Sodium thiocyanate is prepared by heating a mixture of sodium cyanide and sulfur. *See* CYANIDE; SULFUR. [E.E.WR.]

Thioether

One of a group of organosulfur compounds that are also called sulfides, RSR′. The simplest, dimethyl sulfide, $CH_3—S—CH_3$, is obtained in large amounts from sulfite waste liquors of wood treatment and is the precursor of dimethyl sulfoxide, a useful solvent and chemical reactant. Some amino acids, such as methionine and lanthionine, are also sulfides. Mustard gas, $ClCH_2CH_2SCH_2CH_2Cl$, formed by the reaction of sulfur dichloride, SCl_2, and ethylene, is a well-known vesicant. The thioethers bear a formal resemblance to the oxygen ethers, and may be synthesized by analogous methods, for example, via alkyl halides and sodium mercaptides.

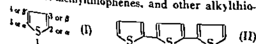


$$RCl + R'SNa \rightarrow RSR' + NaCl$$

Numerous practical uses of sulfides have been claimed, especially as fuel-oil additives, lubricant additives, and agricultural chemicals. *See* AMINO ACIDS; CHEMICAL WARFARE, ETHER, MERCAPTAN, ORGANOSULFUR COMPOUND. [N.K.]

Thiophene

An organic heterocyclic compound containing a diunsaturated ring of four carbon atoms and one sulfur atom. *See* HETEROCYCLIC COMPOUNDS. Thiophene (I), methylthiophenes, and other alkylthio-



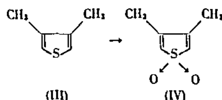
phenes are found in relatively small amounts in coal tar and petrol... Thiophene accompanies benzene in the fractional distillation of coal tar.

...lectively forms ... acid. Alternately, treatment with aluminum chloride selectively polymerizes the thiophene in the benzene to nonvolatile materials. 2,5-Dithienylthiophene (II) has been found in the marigold plant. Biotin, a water-soluble vitamin, is a tetrahydrothiophene derivative.

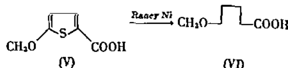
Properties. The parent compound (I) is nearly insoluble in water (forming 0.02–0.04% solutions at 20°), mp −38.2°C, bp 84.2°C, n_D^{20} 1.5287, and specific gravity (20/4) 1.0644. Thiophene has a resonance energy of 29–31 kcal/mole, is stable to heat, and undergoes electrophilic substitutions (nitration, sulfonation, acetylation, halogenation, chloromethylation, and mercuration). Accordingly, thiophene is an aromatic system. Generally, electrophilic substitutions occur with greater ease than with benzene, but less readily than with furan or

... phific agents, ... tion by acid. *See* AROMATIC HYDROCARBON.

Most oxidative processes (nitric acid, ozone, hydrogen peroxide) involving the nucleus have not proved useful in opening the thiophene ring. Peracetic or perbenzoic acid oxidize thiophenes (III) to thiophenesulfones (IV), which behave more as butadiene derivatives than as thiophenes. Sodium



in liquid ammonia and methanol converts thiophene to a mixture of dihydro and acyclic products. Raney nickel strips sulfur from thiophenes in a ring opening reaction, converting (V) to (VI). Catalytic hydrogenation over molybdenum or co-



balt sulfide catalysts at high temperature and pressure, as well as over platinum or palladium catalysts in massive amounts saturates the ring.

Bromine and chlorine react readily with thiophenes, which undergo both substitution and addition reactions. Control of conditions as well as the possibility of dehydrohalogenation by alkali of the products first-formed furnishes halogenated thiophenes in practical preparations. Iodination of thiophene in the presence of *mercuric oxide*, or iodination of mercurated thiophenes gives iodinated derivatives.

Preparation. The thiophene ring system is formed by cyclization of 1,4-dicarbonyl compounds in the

presence of phosphorus sulfides (for example, 2,5-hexadione gives 2,5-dimethylthiophene; 4-oxo-3-ethylpentanoic acid gives 2-methyl-3-ethylthiophene), or by cyclization of hydrocarbons with sulfur or sulfur compounds at elevated temperatures (for example, the reaction of 2-methylbutadiene with sulfur at 320–420° gives 3-methylthiophene; the reaction of ethylbenzene with sulfur in a bimolecular process gives 2,4-diphenylthiophene). The commercial production of thiophene (I) from readily available butane or butadiene awaits only a large-scale demand. A laboratory synthesis converts sodium succinate to thiophene by heating with phosphorus sulfide.

Alkylthiophenes are prepared by ring synthesis, by alkylation of thienylmagnesium halides with sulfate or sulfonate esters, or by reduction of thiophene ketones. 2-Vinylthiophene, potentially of interest as a polymerizable monomer, can be prepared by reducing 2-acetylthiophene to methyl-2-thienylcarbinol, and dehydrating.

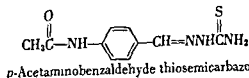
Thiophene aldehydes are prepared by treatment of the thiophene with hexamethylenetetramine (Sommelet process), or with the *N*-methylformanilide–phosphorus oxychloride reagent pair. Friedel-Crafts acylation, often with mild catalysts, gives thiophene ketones in good yields. Thiophene carboxylic acids result from the silver oxide oxidation of thiophene aldehydes, the haloform oxidation of acetylthiophene, and the carbonation of thiophene-metal derivatives. Thiophene aldehydes, ketones, and acids show normal chemical behavior, similar to that of the corresponding benzene derivatives *See* ORGANOSULFUR COMPOUND; THIAZOLE.

[W.J.GE.]

Bibliography: R Adams, et al. (eds.), *Organic Reactions*, vol 6, 1951; H. D Hartough, *Thiophene and Its Derivatives*, 1952.

Thiosemicarbazone

A class of chemical compounds used in the treatment of tuberculosis. These compounds were synthesized in the laboratory of G. Domagk, who had some years earlier introduced the sulfonamides to chemotherapy. It was, in fact, a direct outgrowth of a systematic investigation of the sulfonamides in tuberculosis that led to the finding of the tuberculostatic activity of the thiosemicarbazones. The most prominent member of this group is *para*-acetaminobenzaldehyde thiosemicarbazone (Tibione; Myvizone), which has achieved some clinical success in Europe. In the United States Tibione had a



p-Acetaminobenzaldehyde thiosemicarbazone

brief clinical trial, but it was subsequently dropped because its use was accompanied by a high incidence of serious toxic reactions including anemia, agranulocytosis, and liver and kidney damage. Many modifications of the Tibione structure have been made with little success in producing a more highly active compound with reduced toxicity.

In view of the close similarity between the organism causing tuberculosis and that causing leprosy, the thiosemicarbazones have logically been tried in the treatment of leprosy. Despite some early successes with Tibione, it was found that relapses were common on prolonged treatment. Thus it appears that the thiosemicarbazones are a more promising group for exploitation as antitubercular agents. *See* CHEMOTHERAPY; LEPROSY; TUBERCULOSIS. [N.J.C.]

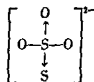
Thiosulfate

A negative ion having the formula $S_2O_3^{2-}$, which is derived from the unstable acid, thiosulfuric acid $H_2S_2O_3$. The usual qualitative test for the thiosulfate ion is to add acid to the substance in question and watch for the white colloidal sulfur and the evolution of sulfur dioxide when the unstable thiosulfuric acid decomposes.



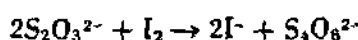
$$2H^+ + S_2O_3^{2-} \rightarrow [H_2S_2O_3] \rightarrow S + H_2O + SO_2$$

Because sodium thiosulfate is formed by heating sodium sulfite and sulfur, this gives a clue to the fact that the two sulfur atoms in the ion have different roles. The ion is actually related to the sulfate ion with the substitution of a sulfur atom for an oxygen, thus:



The central sulfur atom has an oxidation number of 6+, whereas the second has an oxidation number of 2−.

Sodium thiosulfate or hypo is used in photography to "fix" films by dissolving the unreacted silver halide. Sodium thiosulfate reacts quantitatively with iodine solutions in the following manner:

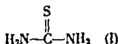


$$2S_2O_3^{2-} + I_2 \rightarrow 2I^- + S_4O_6^{2-}$$

This reaction can be used in volumetric analysis. *See* OXIDATION-REDUCTION; PHOTOGRAPHIC MATERIALS, SULFUR. [E.E.WR.]

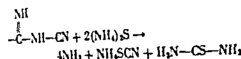
Thiourea

A crystalline, colorless solid (prisms or needles) having formula (I)



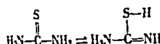
and melting point 180–182°C. Thiourea is relatively insoluble in water (one part in 11) and only slightly soluble in ether. The three most common methods of preparation are (1) heating ammonium thiocyanate, NH_4CNS, to about 180°C where equilibrium with thiourea is established; (2) the action of hydrogen sulfide at about 180°C on calcium

amide; and (3) the reaction of dicyandiamide ammonium sulfide at 60–70°C:

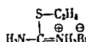


hiourea is the sulfur analog of urea, and like latter, forms addition compounds with hydro-bons (branched chain, alicyclic, and straight-ins of more than 14 atoms). It has been used to tect clothes and furs from insects. *See* UREA.

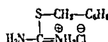
Chemically, thiourea exists in tautomeric forms, (I),



d it often reacts as the latter form, known as eudothiourea. Thus, with an alkyl halide, such as yl bromide, ethyl thiouronium bromide is rmed

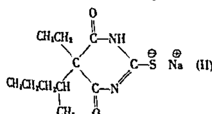


hiouronium salts find use in the preparation of ercaptans, and of alkyl sulfonyl chlorides. Benzyl hiouronium chloride,



is a valuable reagent for the identification of organic acids, because, with the sodium salt of an acid, a crystalline thiouronium salt results.

Condensation of thiourea with substituted malonic esters gives thiobarbituric acid derivatives. Of the thiobarbiturates, 1 methylbutylethylthiobarbiturate as the sodium salt, sodium pentothal (II),



is much used in anesthetic premedication to reduce the amount of general anesthetic needed. *See* CARBOXYLIC ACID; MERCAPTAN, UREID [E B R.]

Thirst

Thirst usually calls to mind the subjective sensations resulting from water deprivation, but studies of the topic have to include the total complex of motivated behavior concerned with the accurate supply of water to the organism. Viewed in this larger context, the study of thirst involves (1) the bodily cues that alert the organism to its need for water, (2) the physiological process that mirrors accurately the quantity of water needed and, of particular interest in this process, (3) the mechanism responsible for the inhibition of drinking when sufficient fluid has been ingested.

Local sign theory. *Serious scientific study of this problem started with an interest in the subjective* sensation of thirst. W. B. Cannon's local sign theory attempted to solve this problem in the areas of both hunger and thirst. *There was exclusive con*-[illegible] *th the localized* peripheral conditions that [illegible]

the mouth and throat *that normally* [illegible] onset of thirst. Briefly, a shortage of water in the organism was said to lower the rate of salivation with the resulting dry mouth then being signaled to the central nervous system *(CNS) via local sensory nerves.* This process, according to the local sign theory, would prompt the organism to drink and *thus replace the needed water. Subsequent studies* have shown that, although a *dry mouth* may very well *be one of the cues serving to remind an organ*ism that water is needed, a much more basic and complex mechanism is involved in the over-all monitoring of drinking behavior. For example, when *salivation is depressed in experimental animals,* they do, indeed, drink more frequently, but over a 24-hour period they still drink no more water in total than normal animals of comparable body *weight. See* HUNGER, NERVOUS SYSTEM.

Water balance concept. The quantitative details of drinking behavior are closely geared to the organism's water balance. This balance is determined *by the algebraic relationship between water intake* and water loss. Water is lost through three routes: (1) perspiration, (2) evaporation from respiratory passages, and (3) urine and feces. Under normal *circumstances terrestrial animals ingest water by* only one route, the drinking of fluids. When intake exceeds loss, the organism is said to be in a state of positive water balance, the reverse condition is *called either negative water balance or the or*ganism is said to be in a state of water deficit. The feeling of thirst sets in when this deficit is on the order of 1% of the body weight. When it *reaches 5–8% of body weight, the organism be*comes weary and approaches collapse. In moving from 10 to 20% deficit, the individual moves from gross physical and mental deterioration toward death. In *experimental animals it has been shown* that the amount of water ingested is controlled by the degree of water deficit, but this still leaves open the question of how the underlying mechanisms operate physiologically.

Physiological mechanism. Animals have been prepared so that water could be introduced directly *into their stomachs* without entering the mouth or throat, and several studies have shed *some light on* the problem of the physiological mechanism involved in thirst. When an adequate amount of water is *placed in the stomach of a thirsty animal,* there is no drinking provided the animal is not allowed access to water for 15–20 minutes after being "pre-

presence of phosphorus sulfides (for example, 2,5-hexadione gives 2,5-dimethylthiophene; 4-oxo-3-ethylpentanoic acid gives 2-methyl-3-ethylthiophene), or by cyclization of hydrocarbons with sulfur or sulfur compounds at elevated temperatures (for example, the reaction of 2-methylbutadiene with sulfur at 320–420° gives 3-methylthiophene; the reaction of ethylbenzene with sulfur in a bimolecular process gives 2,4-diphenylthiophene). The commercial production of thiophene (I) from readily available butane or butadiene awaits only a large-scale demand. A laboratory synthesis converts sodium succinate to thiophene by heating with phosphorus sulfide.

Alkylthiophenes are prepared by ring synthesis, by alkylation of thienylmagnesium halides with sulfate or sulfonate esters, or by reduction of thiophene ketones. 2-Vinylthiophene, potentially of interest as a polymerizable monomer, can be prepared by reducing 2-acetylthiophene to methyl-2-thienylcarbinol, and dehydrating.

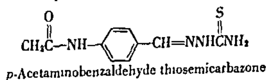
Thiophene aldehydes are prepared by treatment of the thiophene with hexamethylenetetramine (Sommelet process), or with the *N*-methylformanilide–phosphorus oxychloride reagent pair. Friedel-Crafts acylation, often with mild catalysts, gives thiophene ketones in good yields. Thiophene carboxylic acids result from the silver oxide oxidation of thiophene aldehydes, the haloform oxidation of acetylthiophene, and the carbonation of thiophene-metal derivatives. Thiophene aldehydes, ketones, and acids show normal chemical behavior, similar to that of the corresponding benzene derivatives. *See* ORGANOSULFUR COMPOUND; THIAZOLE.

[W.J.CE.]

Bibliography: R. Adams, et al. (eds), *Organic Reactions*, vol. 6, 1951. H. D. Hartough, *Thiophene and Its Derivatives*, 1952.

Thiosemicarbazone

A class of chemical compounds used in the treatment of tuberculosis. These compounds were synthesized in the laboratory of G. Domagk, who had some years earlier introduced the sulfonamides to chemotherapy. It was, in fact, a direct outgrowth of a systematic investigation of the sulfonamides in tuberculosis that led to the finding of the tuberculostatic activity of the thiosemicarbazones. The most prominent member of this group is *para*-acetaminobenzaldehyde thiosemicarbazone (Tibione; Myvizone), which has achieved some clinical success in Europe. In the United States Tibione had a



p-Acetaminobenzaldehyde thiosemicarbazone

brief clinical trial, but it was subsequently dropped because its use was accompanied by a high incidence of serious toxic reactions including anemia, agranulocytosis, and liver and kidney damage. Many modifications of the Tibione structure have been made with little success in producing a more highly active compound with reduced toxicity.

In view of the close similarity between the organism causing tuberculosis and that causing leprosy, the thiosemicarbazones have logically been tried in the treatment of leprosy. Despite some early successes with Tibione, it was found that relapses were common on prolonged treatment. Thus, it appears that the thiosemicarbazones are a more promising group for exploitation as antitubercular agents. *See* CHEMOTHERAPY; LEPROSY; TUBERCULOSIS.

[N.J.G.]

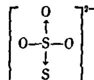
Thiosulfate

A negative ion having the formula $S_2O_3^{2-}$, which is derived from the unstable acid, thiosulfuric acid, $H_2S_2O_3$. The usual qualitative test for the thiosulfate ion is to add acid to the substance in question and watch for the white colloidal sulfur and the evolution of sulfur dioxide when the unstable thiosulfuric acid decomposes.



$$2H^+ + S_2O_3^{2-} \rightarrow [H_2S_2O_3] \rightarrow S + H_2O + SO_2$$

Because sodium thiosulfate is formed by heating sodium sulfite and sulfur, this gives a clue to the fact that the two sulfur atoms in the ion have different roles. The ion is actually related to the sulfate ion with the substitution of a sulfur atom for an oxygen, thus:



The central sulfur atom has an oxidation number of 6+, whereas the second has an oxidation number of 2−.

Sodium thiosulfate or hypo is used in photography to "fix" films by dissolving the unreacted silver halide. Sodium thiosulfate reacts quantitatively with iodine solutions in the following manner:



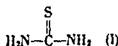
$$2S_2O_3^{2-} + I_2 \rightarrow 2I^- + S_4O_6^{2-}$$

This reaction can be used in volumetric analysis. *See* OXIDATION-REDUCTION; PHOTOGRAPHIC MATERIALS; SULFUR.

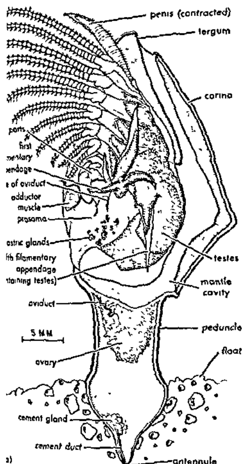
[E.E.W.B.]

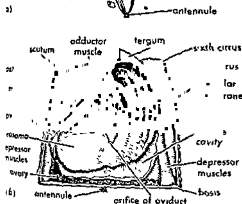
Thiourea

A crystalline, colorless solid (prisms or needles) having formula (I)



and melting point 180–182°C. Thiourea is relatively insoluble in water (one part in 11) and only slightly soluble in ether. The three most common methods of preparation are (1) heating ammonium thiocyanate, NH_4CNS, to about 180°C where equilibrium with thiourea is established; (2) the action of hydrogen sulfide at about 180°C on calcium





g 2 (a) *Lepas fascicularis* Ellis and Solander, with ght side of capitulum, peduncle, and float removed) *Balanus*, with right side of wall removed. (From .W. Hegner, *Invertebrate Zoology*, Macmillan, 1933)

ody is enclosed by an immovable wall formed by ie overlapping of four, six, or eight plates and a iovable operculum of one or two pairs of plates he opercular plates have muscles for closing, and

palps, maxillulae, and maxillae. The digestive system is differentiated into a foregut, enlarged midgut with digestive glands, hindgut, and anus. Paired excretory glands open on the maxillae. Circulation is effected by lacunae, while respiration is mainly by the surfaces of the body and the mantle. The nervous system corresponds to that of other Crustacea but some ganglia are fused.

In the hermaphrodite the reproductive system consists of ovaries with paired oviducts opening on the first thoracic segment, and testes with paired seminal vesicles opening into the penis. Impregnation occurs when spermatozoa are deposited near the openings of the oviducts by the penis of an adjacent barnacle; occasionally self-fertilization occurs. A secretion from the oviduct unites the fertilized eggs into thin laminae, and development to the nauplius, or less commonly, to the cypris stage occurs in the mantle cavity. The number of larvae released at the cypris stage is small (about 50) compared to the number released at the nauplius stage (100,000 or more). Fertilization may occur two to three weeks after settlement, semi-annually, annually, or infrequently not until the second year.

Immediately after settlement, growth is very rapid (up to 1–2 mm per day) accompanied by frequent molts of the body covering, mantle lining, and, in sessile barnacles, the opercular membrane Rate of growth is greater in the tropics. Duration of life varies from 1 year in many intertidal species to at least 10 years in some subtidal species. See CIRRIPEDIA. [D.P.H.]

Thorianite

A radioactive mineral with the idealized composition ThO_2, thorium dioxide. Rare earths and uranium are often present in variable amounts, together with small amounts of radiogenic lead. Thorianite is isometric in crystallization and usually occurs as worn cubic crystals. The hardness is about 7 on Mohs scale, and the specific gravity is 9.7–9.8. The color is brownish black to reddish brown and the luster usually is resinous Thorianite is a primary mineral found chiefly in pegmatites. It is best known as a detrital mineral associated with thorite, monazite, ilmenite, and other heavy minerals in the black sands of stream and beach deposits It has been obtained commercially from detrital deposits and pegmatites in Madagascar and Ceylon *See* RADIOACTIVE MINERALS; RARE-EARTH ELEMENTS; THORIUM; URANIUM, *see also* LEAD ISOTOPES, GEOCHEMISTRY OF; NUCLEAR FUELS, PEGMATITE [C.FR.]

Bibliography: C. Frondel, *Systematic Mineralogy of Uranium and Thorium*, USGS Bull. 1064, 1958.

Thorite

A mineral, thorium silicate, in which the element thorium was discovered in 1828. Thorite is tetragonal in crystallization and has a crystal structure identical with that of the nesosilicate zircon, $ZrSiO_4$. The idealized chemical formula of thorite

watered." Evidence from these and related experiments suggests that water must enter the circulation of the animal in order to influence the level of thirst.

Two important circulatory changes which result from water deprivation can also be produced by other means. The changes are (1) decreased blood volume and (2) increased hemoconcentration, that is, increased concentration of solid material in the circulating blood. Hemorrhage produces the first circumstance exclusively, while the intravenous injection of strong salt solution leads to the latter. Both conditions cause increased drinking. It seems that both of these humoral changes produce thirst by virtue of a resultant effect common to both, namely, cellular dehydration.

Cellular dehydration results when water moves from body tissues into the blood. Although the tissues of the body in general are involved in this loss of water to the circulation, the critical cellular dehydration as far as thirst is concerned is thought to take place in certain cells in the hypothalamus in the brain. These cells may be the same ones that are involved in triggering the release of antidiuretic hormone by the pituitary, in which connection they have been referred to as osmoreceptors (*see* HORMONE, NEUROHYPOPHYSEAL). Exactly how the neural activity in these hypothalamic cells might lead to feelings of thirst or the ingestion of water is not known. It has been shown, however, that the injection of salt solution directly into this area of the hypothalamus does produce immediate drinking behavior in experimental animals which are not in a state of water deficit. *See* BRAIN.

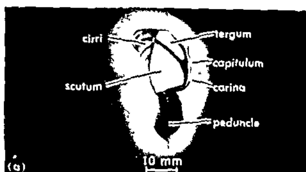
The cessation of a period of drinking, under normal circumstances, seemingly comes about too rapidly for absorption of the ingested fluids to play a role in stopping the animal's immediate drinking response. There is experimental evidence to suggest that this step in the complex process of drinking is quantitatively monitored by the swallowing mechanism itself. Sensory cues resulting from stomach distention may also be involved.

The question has been raised as to whether the antidiuretic hormone, which reduces water loss through the kidneys, is directly involved in thirst and drinking behavior. The influence of this hormone in reducing water loss via the kidney, when a water deficit exists, makes it clear that it is at least influential with regard to the organism's water balance. There is no convincing evidence, however, to suggest a more direct role in the regulation of drinking. *See* BODY RHYTHM. [R.A M.]

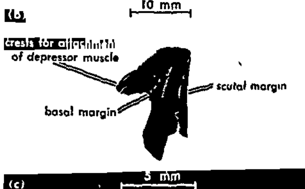
Bibliography: C. T. Morgan and E Stellar, *Physiological Psychology*, 2d ed., 1950.

Thoracica

An order of the subclass Cirripedia. These crustaceans are permanently attached in the adult stage. The mantle is usually protected by calcareous plates, and six pairs of biramous thoracic appendages are present. The abdomen is lacking or represented by caudal furca. Antennules are discernible







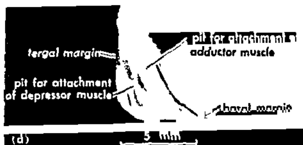


Fig. 1. (a) Lepas anatifera L (from D P Henry, The Cirripedia of Puget Sound with a key to the species, Univ Wash Publs. Oceanog., 4(1) 1–48, 1940). (b) Balanus eburneus Gould, lateral view of shell. (c) *B. eburneus*, inner view of tergum. (d) *B eburneus*, inner view of scutum (from D P Henry, American waters, *Friday Harbor Symposium in Marine Biology*, University of Washington Press, 1959).

in the adult; cement glands are strongly developed and these organisms are usually hermaphroditic.

There are three suborders, the Lepadomorpha which are stalked barnacles, and the Verrucomorpha and Balanomorpha, both sessile. Stalked barnacles are attached by a peduncle and the animal's body is enclosed in a soft mantle (capitulum), usually protected by calcareous plates. The cirri are protruded through the ventral opening by movements of the body and the opening is closed by adductor muscles. In sessile barnacles the animal's

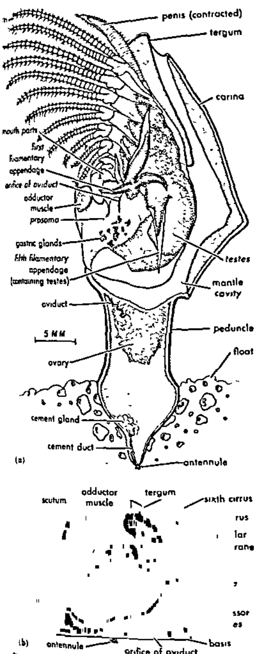


Fig. 2. (a) *Lepas fascicularis* Ellis and Solander, with right side of capitulum, peduncle, and float removed (b) *Balanus*, with right side of wall removed (*From R. W. Hegner, Invertebrate Zoology, Macmillan, 1933*)

body is enclosed by an immovable wall formed by the overlapping of four, six, or eight plates and a movable operculum of one or two pairs of plates The opercular plates have muscles for closing, and in the Balanomorpha also for opening Sessile barnacles have a relatively wide membranous or calcareous basis for attachment.

Barnacles use their spinose cirri to capture food

palps, maxillulae, and maxillae. The digestive system is differentiated into a foregut, enlarged midgut with digestive glands, hindgut, and anus. Paired excretory glands open on the maxillae. Circulation is effected by lacunae, while respiration is mainly by the surfaces of the body and the mantle. The nervous system corresponds to that of other Crustacea but some ganglia are fused.

In the hermaphrodite the reproductive system consists of ovaries with paired oviducts opening on the first thoracic segment, and testes with paired seminal vesicles opening into the penis. Impregnation occurs when spermatozoa are deposited near the openings of the oviducts by the penis of an adjacent barnacle; occasionally self-fertilization occurs. A secretion from the oviduct unites the fertilized eggs into thin laminae, and development to the nauplius, or less commonly, to the cypris stage occurs in the mantle cavity. The number of larvae released at the cypris stage is small (about 50) compared to the number released at the nauplius

Immediately after settlement, growth is very rapid (up to 1–2 mm per day) accompanied by frequent molts of the body covering, mantle lining, and, in sessile barnacles, the opercular membrane. Rate of growth is greater in the tropics. Duration of life varies from 1 year in many intertidal species to at least 10 years in some subtidal species. *See* CIRRIPEDIA [D.P.H.]

Thorianite

A radioactive mineral with the idealized composition ThO_2, thorium dioxide. Rare earths and uranium are often present in variable amounts, together with small amounts of radiogenic lead. Thorianite is isometric in crystallization and usually occurs as worn cubic crystals. The hardness is about 7 on Mohs scale, and the specific gravity is 9.7–9.8 The color is brownish black to reddish brown and the luster usually is resinous Thorianite is a primary mineral found chiefly in pegmatites. It is best known as a detrital mineral associated with thorite, monazite, ilmenite, and other heavy minerals in the black sands of stream and beach deposits It has been obtained commercially from detrital deposits and pegmatites in Madagascar and Ceylon *See* RADIOACTIVE MINERALS; RARE-EARTH ELEMENTS; THORIUM; URANIUM; *see also* LEAD ISOTOPES, GEOCHEMISTRY OF; NUCLEAR FUELS; PEGMATITE [C FR.]

Bibliography: C Frondel, *Systematic Mineralogy of Uranium and Thorium*, USGS Bull. 1064, 1958

Thorite

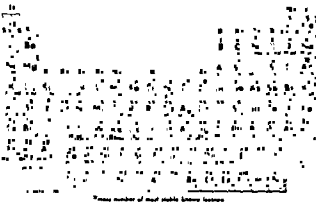
A mineral, thorium silicate, in which the element thorium was discovered in 1828 Thorite is tetragonal in crystallization and has a crystal structure identical with that of the nesosilicate zircon, $ZrSiO_4$ The idealized chemical formula of thorite

is $ThSiO_4$. All natural material departs widely from this composition owing to the partial substitution of uranium, rare earths, calcium and iron for thorium. Structurally, thorite usually has completely lost its crystallinity because of radiation damage from the contained uranium and thorium (metamict state). The specific gravity ranges between about 4.3 and 5.4. The hardness is about 4½. The color commonly is brownish yellow to brownish black and black.

Thorite occurs chiefly in pegmatites. It also occurs as an accessory mineral in black sands and other detrital deposits derived from granitic or gneissic terranes. *See* RADIOACTIVE MINERALS; THORIUM. [C.FR.]

Thorium

Chemical element number 90, thorium, Th, was discovered by J. J. Berzelius in 1828. However, little use was found for thorium before the development



of the incandescent gas mantle by C. A. von Welsbach in 1885. Several thousand pounds of thorium oxide still go into the annual production of these mantles. A small amount of thorium oxide is incorporated in the tungsten that is used for electric light filaments. The oxide is employed in catalysts for the promotion of certain organic chemical reactions. Thorium oxide has special uses as a high-temperature ceramic material. The metal or its oxide is employed in some electronic tubes, photocells, and special welding electrodes. The metal can serve as a getter in vacuum systems and in gas purification and as a scavenger in some metals.

Because of its high density, chemical reactivity, mediocre mechanical properties, and relatively high cost, thorium metal has no market value as a structural material; however, it has important applications as an alloying agent in some structural metals. Perhaps the major use for thorium metal, outside the nuclear field, is in magnesium technology. Approximately 3% thorium, added as an alloying ingredient, imparts to magnesium metal high-strength properties and creep resistance at elevated temperatures. The magnesium alloys containing thorium, because of their light weight and desirable strength properties, are being used in aircraft engines and in airframe construction.

Thorium can be converted in a nuclear rea[ctor] to uranium-233, an atomic fuel. The system of [tho]rium and uranium-233 gives promise of comp[lete] utilization of all thorium in the production [of] atomic power. The energy available from [the] world's supply of thorium has been estimated [as] greater than the energy available from all of [illegible]

[illegible]

and commercially most important thorium-bear[ing] mineral, is widely distributed in nature. Import[ant] deposits occur along the shores of India, Bra[zil] and Ceylon. Other extensive deposits of mona[zite] are found in South Africa, Russia, Scandina[via] and Australia. Domestic sources include depo[sits] in Florida, Idaho, and the Carolinas. Monazite [is] chiefly obtained as a sand, which is separated fr[om] other sands by physical or mechanical means, f[ol]lowing dredging operations. The monazite sa[nd] concentrate is essentially an orthophosphate [of] rare-earth elements, and generally contains 3–10[%] ThO_2. Other thorium-bearing minerals of les[s] importance include thorite, thorianite, and uran[o]thorite.

Metallurgical extraction. Processes for thoriu[m] recovery generally start by digestion of the mon[a]zite sand with either hot concentrated sulfuric ac[id] or hot concentrated caustic. Subsequent chemic[al] treatments, varying greatly even with the sam[e] initial treatment, yield a concentrate of impu[re] thorium. This impure concentrate may be furth[er] treated by a liquid-liquid extraction process [to] yield high-purity thorium. For a system consistin[g] of water, tributyl phosphate, nitric acid, thoriu[m] and the associated impurities, conditions for opera[tion] of an extractor can be set up to remove th[e] thorium with the water-immiscible tributyl pho[s]phate phase, while the impurities are carried awa[y] in the aqueous phase. Generally, the purified tho[rium] is either crystallized from solution as the ni[trate] or precipitated as the oxalate. From thes[e] pure salts, the oxide or other compounds of tho[rium] can be prepared.

Because thorium is quite reactive, some difficulty is experienced in preparing thorium metal. Only by electrolysis or by treatment with elements high in the electromotive force series (the alkali and alkaline-earth metals), can good-quality thorium metal be satisfactorily prepared directly from its compounds.

The calcium reduction of ThO_2 has been widely used for many years to prepare thorium metal. In this process, granular calcium metal is mixed with the thorium oxide and charged into a lined iron crucible which is then filled with an inert gas and heated to near 1000°C to form the thorium metal powder and calcium oxide. After cooling to room temperature, the thorium powder is recovered by leaching and then drying. Powder metallurgy techniques are employed to obtain the metal in mass.

The electrodeposition of thorium from a bath, consisting of thorium chlorides or fluorides dis-

solved in fused alkali halides, yields granular thorium that may be pressed and sintered to give massive pieces of ductile metal

Large-scale production of thorium metal is carried out by a bomb process. The charge, consisting of a mixture of thorium tetrafluoride, granular calcium metal, and zinc chloride, is placed in a refractory lined vessel which is closed by a lid. This charged bomb is placed in a furnace held at about 650°C where, after several minutes, the charge ignites spontaneously, and the resulting reaction yields a slag of *calcium fluoride and calcium* chloride and an alloy of thorium and zinc. The temperature reached by the reaction in the charge is sufficient to melt the products, and the thorium-rich alloy collects as a molten pool under the liquid slag. The bomb is allowed to cool, and then the solid piece of thorium alloy is removed and cleaned of adhering slag. Next, the zinc is removed by heating the alloy in a vacuum at a temperature of

by shaping the sponge in the form of bars and melting *these by consumable electrode arc-melting*. Good quality thorium metal can be readily worked *to* shape by standard methods of fabrication See VACUUM METALLURGY.

Properties of the element. Thorium has an atomic weight of 232. The metal has a density of 11 7 g/cm³. Good quality thorium metal is relatively soft and ductile. It can be shaped readily by any of the ordinary metal-forming operations. It must be protected, however, to prevent oxidation in treatments involving high temperatures. The massive metal is silvery in color, but it tarnishes on long exposure to the atmosphere; finely divided thorium has a tendency *to be pyrophoric in air*.

The atoms of thorium in the metal are arranged in a face-centered *cubic system at all* temperatures

Thorium is a member of the actinide series of elements, which includes protactinium, uranium, and the synthetic transuranic elements. It is radioactive with a half life of about 14×10^9 years It is the first member of the radioactive decay series which in a chain of 10 successive disintegrations (α and β combined) finally terminates as lead-208

All of the nonmetallic elements, except the rare gases, form binary compounds with *thorium* Binary intermetallic compounds have been reported for thorium with beryllium, *magnesium. boron*, aluminum, and silicon, and with most of the heavier metals in each of the long *periods of the* periodic table, beginning with the metals in group VIIa A number of the *intermetallic* compounds of thorium, especially those with copper, silver, and gold, are quite pyrophoric. A study of the binary alloy systems of most of the metals in the subgroups a of *groups* III, IV, V, and VI with thorium metal has shown no evidence of intermetallic compounds in these systems.

Principal compounds. Thorium does not impart any visible spectrum colors *to its inorganic compounds* or their solutions. With minor exceptions, thorium exhibits a valence of 4+ in all of its salts. Chemically, it has some resemblance to zirconium and hafnium. The most common soluble compound of thorium is the nitrate which, as generally prepared, appears to have the formula $Th(NO_3)_4 \cdot 4H_2O$.

The common oxide of thorium is ThO_2, thoria, which can be *obtained by thermal decomposition* of the nitrate, hydroxide, oxalate, or other compounds of thorium. A peroxide of thorium, Th_2O_7 with water of hydration, and the hydroxide, $Th(OH)_4$, can *be precipitated from solutions of* thorium salts.

The halogens form a variety of salts with thorium Thorium tetrahalides of the general formula ThX_4 (X = halogen), anhydrous and with varying *degrees of hydration, are known.* $ThOX_2$ *and* $Th(OH)_2X_2$ with and without water of hydration are known. The *halides of thorium also tend to* form double salts with other halides, such as those of the alkali metals.

Thorium sulfate can be obtained in the anhydrous form, or with 2, 4, 6, 8, or 9 molecules of water of crystallization. A somewhat insoluble basic sulfate forms when a dilute water solution of thorium sulfate is boiled. Double sulfates of thorium and alkali metals or ammonium are known. The hydrosulfate and the thiosulfate of *thorium are* water-insoluble compounds.

Thorium carbonates, phosphates, iodates, chlorates, chromates, molybdates, and other inorganic salts of thorium are well known. Thorium also forms salts with many organic acids, of which the water-insoluble oxalate, $Th(C_2O_4)_2 \cdot 6H_2O$, is fairly important in the preparation of pure compounds of thorium.

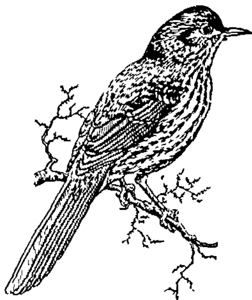
Analytical methods. Thorium in small quantities in rocks and other natural sources can be estimated by a study of the radioactivity of the sample. The chemical analysis of materials for thorium, however, generally involves getting the *thorium into* solution with sulfuric acid The thorium must then be isolated from other *interfering ions, and this* may involve obtaining from the sulfate solution a *precipitate such as thorium iodate* or pyrophosphate which may be subsequently treated for *gravimetric, titrimetric, or colorimetric* determination of the thorium. The gravimetric method generally depends on formation of the oxalate, which is subsequently calcined to ThO_2 and weighed. A titrimetric determination can be made on a thorium solution by titration with ammonium molybdate, which forms a precipitate with thorium; an outside indicator may be employed to determine the end

point of this titration. In a colorimetric method, a complex organic compound, referred to as Thorin, gives, with thorium ion, a color that can be measured to indicate the quantity of thorium present. *See* ACTINIDE ELEMENTS; RADIOACTIVITY. [H.A.W.]

Bibliography: G. T. Seaborg and J. J. Katz (eds.), *The Actinide Elements*, NNES, Div. IV, vol. 14A, 1954; G. T. Seaborg and L. I. Katzin (eds.), *Production and Separation of U-233*, USAEC Report TID-5222, 1951; H. A. Wilhelm (ed.), *The Metal Thorium*, 1958.

Thrasher

Any of 10 species of the genus *Toxostoma*, of the perching bird family Mimidae, 7 of which occur in the United States, 6 of them in the far West or Southwest. The brown thrasher, *T. rufum*, is the best known. It ranges from Alberta to Texas and east to the Atlantic coast. This long-tailed, curve-



The brown thrasher, *Toxostoma rufum*; length to 12 in. (*From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

billed songbird is a familiar species in the yard and garden, where it is often incorrectly called the brown thrush. Its song, usually delivered from a tree top, is rich and varied, similar to that of its relative, the mockingbird. *See* CATBIRD, MOCKINGBIRD; PASSERIFORMES. [J.D.B.]

Threading

The forming of a ridge and valley of uniform cross section which spiral about the inner or outer diameter of a cylinder or cone in an even and continuing manner. The work must be produced with sufficient uniformity and accuracy so that the resulting threaded part will accomplish its intended purpose of fastening, transmitting motion or power, or measuring.

Screw threads are classed as either external or internal. Thread chasing by cutting dies and single-point turning plus milling, hobbing, grinding, and rolling are the usual means of producing ex-

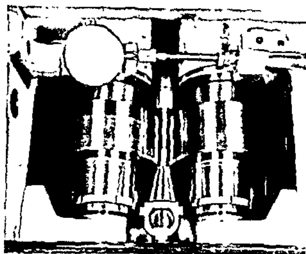


Fig. 1. *Thread rolling machine set up to thread* steel aircraft jet engine mount, with minimum hardne[...] of 40 Rockwell C. Work piece rotates at 1320 rp[...] thread rolling dies contact the work for only 0.24 se[...] producing nearly perfect concentricity and a pit[...] variation of less than 0.0005 in. (*Landis Machine Co*[...]

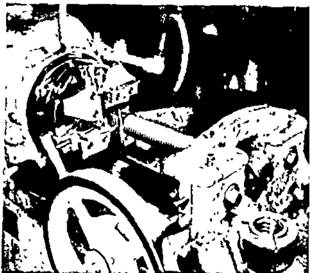
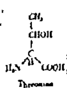


Fig. 2. Threading machine with heat treated die head set up to cut 3 pitch extra deep thread at 7.6 surface feet per minute in adjusting screw for construction machine. (*Landis Machine Co.*)

ternal threads. Most internal threading is done by tapping plus internal single-point turning, hobbing, grinding, and milling.

Although machine threading is done on various machine tools, special threading machines may be used for quantity production. *See* MACHINING OPERATIONS; SCREW FASTENER; SCREW THREADS. [A.T.]

Threonine



Physical constants of the L isomer at 25°C
$pK_1(COOH)$ 2.71 $pK_2(NH_3^+)$ 9.62
Isoelectric point 6.16
Optical rotation $[\alpha]_D(H_2O)$ −28.5 $[\alpha]_D(5\ N\ HCl)$ −15.0
Solubility (g/100 ml H_2O) 20.5 (0°C)

An amino acid which is considered essential for normal growth of animals. The amino acids are

characterized physically by the following: (1) the pK, or the dissociation constant of the various titratable groups; (2) the isoelectric point or pH at which a dipolar ion does not migrate in an electric field; (3) the optical rotation or the rotation im- ... polarized light (fre- ... passing ... grams in ... 100 milliliters; (4) solubility. *See* EQUILIBRIUM, IONIC, ISOELECTRIC POINT; OPTICAL ACTIVITY; SPECTROPHOTOMETRIC ANALYSIS.

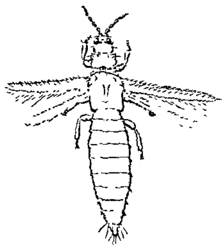
Threonine reacts with periodate to form glyoxylate, ammonia, and acetaldehyde, and is a biosynthetic precursor of isoleucine in microorganisms. Threonine is biosynthesized from aspartic acid (see AMINO ACIDS).

The major pathway in metabolic degradation starts with the nonoxidative deamination to α-ketobutyric acid. This keto acid may be oxidatively decarboxylated to propionyl-CoA, or transaminated to form α-aminobutyric acid. Another pathway of threonine degradation involves an initial cleavage to glycine and acetaldehyde. [E.A.AD.]

Thrip

Any insect belonging to the order Thysanoptera. Thrips are usually very small insects, the American species being 5 mm or less in length; however, some tropical species are almost ½ in long. They have two pairs of similar wings that are narrow and fringed all around with long hairs. Thrips have mouthparts of the sucking type. There are about 3200 widely distributed species. Parthenogenesis is common, and males are unknown in some species.

Thrips sometimes occur in great numbers and several species are harmful. A few bite man. Various species are major pests on such crops as gladioluses, onions, tobacco, and tomatoes. Most species feed on plants, but some eat fungus spores, and a few are predaceous on other insects. *See* INSECTA; THYSANOPTERA. [J.D.B.]



The flower thrip, *Frankliniella tritici*; length about 1/25 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Throat

The region that includes the pharynx, the larynx, and related structures. Both the nasal passages and the oral cavity open into the pharynx, which also contains the openings of the Eustachian tubes from each ear. The lower portion of the pharynx connects with the esophagus and the trachea, or windpipe. The pharynx is shaped like a stocking top which is suspended from the base of the skull and the jaws; its walls are surrounded by three constrictor muscles that primarily aid swallowing.

The larynx, or voice box, is marked externally by the shield-shaped thyroid cartilages which meet in the center to form the Adam's apple. The larynx contains the vocal cords that act as sphincters for air regulation and, in man, permit phonation. The lower end of the larynx is continuous with the trachea, a tube composed of cartilaginous rings and supporting tissues.

The term throat is also used in a general sense to denote the front of the neck. *See* LARYNX, PHARYNX. [E.G.ST.]

Thrombosis

The formation of a thrombus. A thrombus is a solid body formed during life, composed of the elements of the blood—platelets, fibrin, red cells, and leukocytes.

Thrombus formation. Thrombosis is essentially platelet deposition and may occur on any blood vessel wall where the endothelium is damaged. Since platelets release thromboplastinogen, which in turn activates the clotting mechanism, clotting and thrombosis may occur together. Platelets exhibit a tendency to stick together. This tendency is increased if their number is increased, as in thrombocytosis, if the velocity of the stream is decreased below a certain speed, or if the endothelium is roughened. The same factors promote thrombus formation. Heparin, however, causes the platelets to lose their stickiness. The flow of blood is slower in the veins, which may be one reason for the greater frequency of thrombi there. Other determinative factors, as yet unknown, must also come into play.

As the platelets fall out of the blood stream they tend to fall like grains of sand in a moving stream, covering any roughness in the wall. A series of ridges, or laminae, known at the lines of Zahn then forms at right angles to the stream. Leukocytes then stick to these. With the activation of thromboplastin, the clotting mechanism is set in motion and fibrin is added to the thrombus. The thrombus is thus composed of platelets, fibrin, and a mixture of blood cells.

With occlusion of a vessel, true thrombosis ceases, but clotting continues. The clot may be propagated to the next side branch. The propagated clot may float freely in the vessel, anchored only by the thrombus at the base. Under these conditions a piece may break off and travel in the blood stream, forming an embolus.

Blood clotting. This is the process which involves the change of blood from a fluid state to a solid clot. It may be said that three phases are involved in blood clotting. (1) Thromboplastin formation is activated by a thromboplastin precursor, thromboplastinogen, which is released with the disintegration of platelets. (2) Thromboplastin reacts with several factors in the presence of calcium salts to convert the prothrombin of the plasma into thrombin. (3) The thrombin converts the fibrinogen of the blood into fibrin by polymerizing the molecules to form long fibrils, which constitute the resulting fibrin clot.

In summary, a soluble protein, fibrinogen, is converted into an insoluble protein, fibrin, by means of an enzyme, thrombin. Other factors, such as factor V and factor VII (the heat labile and stabile factors), are involved in the conversion of prothrombin into thrombin, and other factors enhance the thromboplastic activity of the platelets. *See* BLOOD; SERUM.

Location of thrombi. Thrombi may form in the veins, arteries, capillaries, or within the heart itself. The most common sites for the formation of thrombi are in the veins. Venous thromboses can be divided into two groups, venous thrombosis or phlebothrombosis, and thrombophlebitis.

Phlebothrombosis. Venous thrombosis, or phlebothrombosis, is the most important type and is often a sequela of generalized circulatory failure, trauma, or prolonged bed rest from any cause. When this type involves the great veins to a large extent, it may be followed by fatal pulmonary embolism. A common initial site for the formation of this group is the deep veins of the leg. *See* EMBOLISM.

Thrombophlebitis Thrombophlebitis is the formation of a thrombus secondary to infection or inflammation of the wall of the vein. In this case the thrombus is usually firmly attached to the vessel wall; hence the chance for embolization is much less. There is, however, the danger that the infected thrombus may break up, with resulting septic emboli. With the advent of asepsis and antibiotics this type of thrombosis has become less important.



Organized thrombus with reconalization; C, new channels.



Organized thrombus with scar formation; C, new channels.

Arterial thrombosis is usually due to local causes. A common predisposing lesion of the vessel wall is arteriosclerosis. This results in roughening of the intima and initiation of the thrombotic process. A common site for this process is the coronary arteries. Thrombosis of a major branch of the coronary tree may be followed by a myocardial infarct if the collateral circulation to the myocardium is not adequate. A similar event may occur in the cerebral arteries with a resulting cerebral infarct. *See* INFARCTION.

Thrombi are found in the heart in three main sites: (1) on inflamed or extremely narrowed valves, forming vegetations; (2) on a necrotic endocardial surface, as is associated with a myocardial infarct; and (3) in stagnation of blood in the auricle, as in auricular fibrillation.

So-called capillary thrombi are really hyalin masses formed by the fusion of red blood cells.

Clotting after death Following death the blood clots in the heart, arteries, and veins, but not in the capillaries. Postmortem clotted blood in the large vessels contains all of the cellular elements and is soft and red. This is the so-called currant jelly clot. When this clotting is slow, however, the red cells fall to the bottom and the clot consists of leukocytes and fibrin. This type of clot has a pale yellow color, is firmer in consistency, and is known as a chicken fat clot The clot microscopically has a uniform homogeneous appearance and shows none of the characteristic architecture of a thrombus.

Fate of thrombi. The fate of thrombi once they are formed, is limited If the thrombus is infected it may break up, the pieces carrying the infected material to another site. The fibrin within the thrombus may contract, causing the thrombus to retract from the vessel wall. This newly formed space may then become endothelized to form a new channel The thrombus can become completely absorbed consequent to the activity of the leukocytes. Materials with enzymatic activity have been used clinically in an attempt to dissolve newly formed thrombi, and have met with limited success.

Finally, a process of organization may ensue. From the lining of the vessel, capillary buds grow into the thrombus, gradually absorb the mass, and replace it with organization tissue, as in any re-

...ative process. During the process, one or more ...scular channels can be formed resulting in a partial reestablishment of the patency of the lumen. This is called canalization of a thrombus. In other instances the lumen is completely obliterated by scar. If this scar tissue becomes calcified it forms a phlebolith. This process occurs most often in the ... of the leg. See INFLAMMATION.

...nal thrombosis can result in an infarct of the region supplied by the vessel if the collateral circulation is inadequate. With mesenteric thrombosis, gangrene of the bowel may ensue. Under these circumstances surgical intervention must follow promptly or death will result.

Occlusion of any vein causes stasis of blood in the region, with an acute passive congestion of the region being drained. Thrombosis of varices of the rectum or lower extremities may result in local necrosis and ulceration. Finally, a thrombus may become detached from the vessel wall and travel in the blood stream as an embolus. [R.A.V.]

Bibliography: W. A. D. Anderson (ed.), *Pathology*, 3d ed., 1957; W. Boyd, *A Text-book of Pathology*, 6th ed., 1953.

Throttled flow

Flow which is forced to pass through a restricted area, where the velocity must be increased. Most of the kinetic energy produced by reduction of pressure in passing through the constriction is generally converted into thermal energy by turbulent eddying. The net result is a loss in mechanical energy in the system. When a gas is throttled, as by a globe valve, the velocity a short distance downstream from the valve is only a little higher than before the throttling section in most cases, the ...

... the amount of flow may be controlled.

A special throttling effect is produced when gas flows through a constriction, such as a nozzle, at sonic velocity. When this occurs, further reduction of downstream pressure does not alter upstream conditions and the flow remains constant. [V.L.S.]

Thrush

Any of several species of perching birds of the family Turdidae. Most of the birds commonly called thrushes in the United States belong to the genus *Hylocichla*, all five species of which are North American. The genus includes some famous songbirds. Best known to most Americans is the wood thrush, *H. mustelina*, which nests in the eastern deciduous forests. In the northern coniferous forests *H. guttata*, the hermit thrush, is a favorite singer. The veery, or willow thrush, *H. fuscescens*, breeds in the moist, mixed forests across the northern United States and south in the mountains into Colorado. All three of these, as well as the others



The wood thrush, Hylocichla mustelina, length 8½ in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

in the genus, have olive-brown backs and spotted or streaked breasts, which makes specific identification difficult.

The varied thrush, *Ixoreus naevius*, lives in western North America, from Alaska to California. It is more like the robin than the typical thrushes, being distinguished from the latter by a dark band across the chest, which is obscure or lacking on the brownish females. See PASSERIFORMES. [J.D.B.]

Thrust

In aviation context, thrust is the force that propels an aircraft or a missile. It is usually expressed in pounds (English gravitational system) or in kilograms (metric gravitational system). It can be expressed in poundals (English absolute system) or dynes (metric absolute system). Conversion between these units is given in the table. The power

Unit	Pound	Poundal	Kilogram	Dyne
1 pound	1	32.2	0.454	4.45×10^{5}
1 poundal	0.0311	1	0.0141	1.38×10^{4}
1 kilogram	2.20	71.0	1	9.81×10^{5}
1 dyne	2.25×10^{-6}	7.23×10^{-5}	1.02×10^{-6}	1

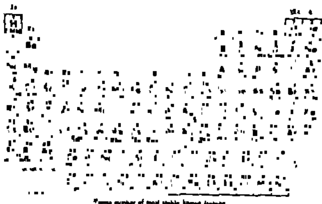
of turbojets, ramjets and rockets is usually ex-

...

work supplied by the power plant. Power in ft-lb/sec can be converted to horsepower by dividing by 550. [R.R.H.]

Thulium

Element number 69, thulium, Tm, is a rare metallic element belonging to the rare-earth group. Its atomic weight is 168.94, and the stable isotope Tm^{169} makes up 100% of the naturally occurring element. It was discovered by P. T. Cleve in 1878. The salts of thulium possess a pale green color and the solutions have a slight greenish tint. The metal is produced by the reduction of the anhydrous



fluoride with calcium or by the vacuum distillation of thulium from a mixture of lanthanum metal and thulium oxide. It has a high vapor pressure at the melting point. When thulium-169 is irradiated in a nuclear reactor, Tm^{170} (half-life 129 days) is formed. The isotope then emits strongly an 84-kev x-ray, and this material is useful in making small portable x-ray units for medical use No electrical equipment is required and the unit has to be recharged with a reactivated thulium button only every few months. The metal becomes antiferromagnetic at low temperatures *See* RARE-EARTH ELEMENTS. [F.H.SP.]

Thunder

When lightning occurs, the resulting expansion of gaseous media is similar to an explosion and produces an atmospheric compression wave which propagates outward at the speed of sound The sound thus produced is thunder. The surge of current down the discharge channel frequently exceeds 30,000 amp. This extremely high current causes a sudden intense heating along the path and leads to the dissociation of the gases in the channel.

In general the sound cannot be heard more than 5–6 miles from the source because of the dispersion caused by the atmosphere. However when conditions between the thunderstorm and the observer are stable, thunder may be heard at a distance of perhaps 10–15 miles. Since lightning is seen almost instantaneously (speed of light = 186,000 miles/sec) while sound travels relatively slowly, one can estimate the distance to the lightning path by multiplying the number of seconds elapsed by 1000 ft/sec, an approximate value of the speed of sound.

The rumbling of thunder results mainly because the twisting nature of lightning paths causes variation in the distance to the sound sources and because multiple strokes in the same path produce a number of compression waves which interfere with one another.

Just before the loud report caused by a nearby lightning strike, a short click is usually heard. This sound has been ascribed to the stepped leader. *See* LIGHTNING. [L.J.B.]

Thunderstorm

A convective storm accompanied by lightning and thunder, rain or rarely snow showers, and often hail. Gusty squall winds are observed near the onset of precipitation. The characteristic cloud is the cumulonimbus, a towering turbulent cloud often having an anvil-shaped top. Sheets or fragments of heavy cirrus, altocumulus, and low stratus or stratocumulus, are often found in association.

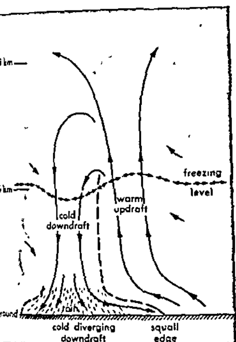
In many regions, particularly the tropics, thunderstorms furnish much or most of the total annual rainfall. Cloudbursts, squall winds, hail, and lightning associated with thunderstorms do many millions of dollars damage annually.

Although much heavier, thunderstorm rain is localized compared with the frontal rain in a cyclone. Thunderstorms may be chaotically scattered over a wide area or may appear as large clusters, often arranged in lines (*see* SQUALL). An isolated thunderstorm is perhaps 3–5 miles across; clusters 20–50 miles across are common, while lines of clusters may be several hundred miles in length. Bases of thunderstorm clouds range from 1000 to 3000 ft in moist air masses, up to 10,000–15,000 ft in dry air. Tops commonly reach 30,000–50,000 ft, but are higher in the tropics and lower in high latitudes.

A thunderstorm consists of one or more cells, identified by distinct patterns of vertical motion, and it goes through characteristic stages of development. In the cumulus or growing stage, a thunderstorm consists of updrafts, or vigorously rising columns of air; in the mature stage (see illustration) both updrafts and pronounced downdrafts are found, and in the dissipating stage gentle downdrafts characterize the whole cloud or cell. Vertical speeds in the drafts have median values near 8 m/sec, but may be several times stronger

Conditions for formation. If unsaturated air is lifted without addition or loss of heat, it cools by expansion at the dry adiabatic lapse rate of 9.8°C/km If saturated it cools at the lesser moist adiabatic lapse rate because released latent heat of condensation partly counteracts the expansional cooling. If the existing decrease of temperature with height outside the cloud is greater than the moist adiabatic rate, a rising saturated air parcel becomes warmer and less dense than its surroundings, and thermodynamic instability exists because the parcel would then be accelerated upward because of its relative buoyancy *See* ATMOSPHERIC ADIABATIC CHANGE

For such instability to be realized, the air rising from lower levels must have high moisture content



agram showing simplified circulation in a vertical ction through a mature thunderstorm. (After G A ckdorff, H R. Byers and R. R. Braham, Jr.)

itiation of convection can result from surface eating, which causes local bubbles or columns of ir to rise to saturation level, or from mechanical lting by a front or over rising ground

Moisture for cloud and rain formation is conensed in the updrafts, because rising air as it ools can hold less water in vapor form. Once rain egins, downdrafts form in the heavy rain area. hese are colder than the cloud environment, and ink by gravity force Coldness of the downdraft is lue to partial evaporation of raindrops, causing

riginated in the initial updraft.

On approaching the earth's surface, the downdrafts are forced to spread out laterally. Gusty squall winds, often of considerable force, are found

by the cold air spreading out from a mature thunderstorm. By this process a cluster of thunderstorm cells may be continuously regenerated even though some cells dissipate.

Because the ground and lower layers of the air are warmest in midafternoon, instability is most pronounced then, and thunderstorm incidence shows a decided peak at that time. Thunderstorms are most common in the tropics, and over middle latitude land areas in summer, but occur occasionally even at high latitudes. In the United States, most frequent occurrences are in the mountainous regions of the Southwest and over the Florida peninsula. Most thunderstorm activity is found in regions of general low-level convergence and weak upward motions, such as on the advancing sides of cyclones and in the intertropical convergence (*see* STORM).

Electrical phenomena. *Thunderclouds* carry mainly positive charges in upper levels and negative charges in lower levels. Theories for charge generation include capturing of ions by droplets, charge separation by breaking of waterdrops, and generation on freezing of supercooled water. The cloud-earth electrical potential before discharge is of order 10^9 volts; in an average lightning stroke about 25 coulombs of electricity are discharged in a current of 25,000 amperes. Thunder results from compression waves set up by violent expansion of air heated in the path of a lightning stroke *See* ATMOSPHERIC ELECTRICITY. [C.W.N.]

Bibliography: T. H. Malone (ed.), *Compendium of Meteorology*, 1951; U.S. Weather Bureau, *The Thunderstorm*, 1949.

Thymus gland

A lymphoid organ in the neck or upper thorax of all vertebrates from elasmobranchs to mammals Its occurrence in cyclostomes is questionable. Embryologically, it arises as an outgrowth from the pharyngeal portion of the alimentary canal. A prominent organ during early life, it commences to regress at the onset of sexual maturity and becomes *largely replaced with fat or fibrous tissue. Despite* intensive investigation, no definite function has yet been determined for it. It is not necessary for life. The thymus is the only organ of considerable size whose place in the bodily economy remains so completely obscure.

Comparative anatomy and embryology. Throughout life in most fishes and amphibians, and during development in other forms, the thymus is closely associated with the epithelial lining (endoderm) of the gill cavities, or the corresponding pharyngeal pouches of nonaquatic forms.

In fishes One thymic anlage, the first accumula*tion of cells in an embryo recognizable* as the rudiment of a developing part or organ, tends to arise from some portion, usually dorsal, of the lining of each pharyngeal pouch Certain of these, however, may remain rudimentary. In some forms the anlagen which develop fuse to form a single strand in each side of the neck The definitive organs may remain closely attached to the epithelium of the gill chambers or withdraw to varying depths in the overlying connective tissue, but they seldom migrate to either lower or higher body levels.

In elasmobranch fishes, the sharks and rays, the thymus appears as slightly lobulated masses of cells. *These lie immediately* beneath the epithelial lining of the dorsal portion of the gill cavity. In some ganoid and teleostean fish, the thymus remains merely a thickened portion of this epithelial layer, so that it resembles the tonsils of amphibians and mammals. These fish thymuses can frequently

be seen as small, whitish spots shining through the mucous membrane of the gill cavity. In fishes as in all vertebrates the thymus must be sought in young animals, since it regresses following sexual maturity.

Thymus in amphibians. The thymus occupies essentially the same position as in fishes. It is dorsal to the larval gill chamber, but somewhat farther removed from the mucous membrane. Being closer to the skin covering the surface of the neck (immediately behind the ear in *Triton taeniatus*), it is readily accessible to surgery.

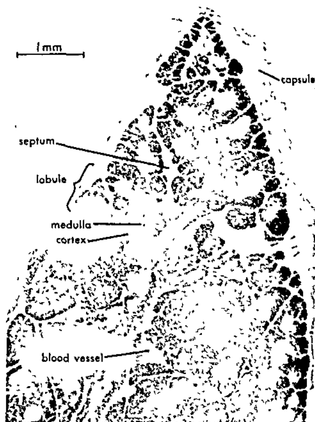
The amphibian anlagen also arise from the dorsal portions of the pharyngeal pouches. Those from the more rostral pouches, I & II, usually fail to develop (*see* PHARYNX). Those which develop may fuse into a single bilateral organ, or there may be a distinctly separate cranial and caudal pair as in *Necturus maculosus*.

Thymuses of reptiles and birds. These show much variation However, they usually occur as two or more pairs of organs located essentially as they are in fishes and amphibians, that is dorsolateral to the esophagus and medial to the jugular vein at approximately the level of the thyroid gland, or more rostrally. Commonly, the anlagen arise from pouches II, III, and IV with that from pouch II remaining rudimentary. In later development a single anlage may separate into several masses which correspond roughly to the spaces between the segmental nerves of the neck.

Mammalian thymus. The usual structure is associated only with the more caudal pharyngeal pouches, III and IV During development, it may migrate yet farther caudally and ventromedially until in man, for instance, the bilateral halves unite to form a single bilobed organ in the ventral wall of the thorax between the pericardial cavity and the sternum. J. A Hammar divides mammals into three groups: those having only thoracic thymuses, those having only cervical thymuses, and those having both thoracic and cervical thymuses. In addition to these chief organs, inconstant accessory thymuses may occur in various other locations

Histology. The thymus is a highly lobulated cellular mass weighing 30–40 g at its point of greatest development in man (11th to 13th year). It is provided with a well-developed connective tissue capsule and interlobular septa, but no lumen. Each highly branched lobule connects to a central core which is an elongated strand only 1–3 mm in diameter. Actually there are two of these cores, representing the original paired thymic diverticula from the third pharyngeal pouches. Thus, the human thymus is organized into two easily separable lobes, each having its own capsule and system of lobules.

The thymic mass is composed chiefly of two types of cells, reticular cells and thymocytes. The *reticular* cells resemble those of other lymphatic organs, lymph node and spleen, in that they have relatively large, pale-staining nuclei and elongated cytoplasmic bodies which tend to interconnect to



Thymus of a 3-month human infant. Histological section through a portion of one lobe

form a spongelike framework within the meshes of which the thymocytes are held.

The thymocytes closely resemble typical lymphocytes of other organs and circulating blood. Thus thymocytes and lymphocytes have the same nuclear, cytoplasmic, and mitochondrial configuration Both react in the same way to irradiation with x-rays and both cell types, when removed to tissue cultures, will transform into plasma cells and macrophages. On the other hand, thymocytes are said to be more sensitive to injected adrenocorticotrophin than are ordinary lymphocytes.

The thymic lobules are separated into two zones, a central medulla, where the lymphocytes (thymocytes) are widely scattered, and a peripheral cortex, in which they are densely massed The strandlike core of each lobe contains only medulla

Hassall's bodies are the only real histological peculiarity of the thymus. These occur in the medulla and are small balls from 30 to over 100 μ in diameter of concentrically layered, flattened cells which are probably reticular. Those at the centers of the bodies tend to degenerate and hyalinize in a manner reminiscent of stratified epithelium and epithelial pearls of other organs.

Histogenesis. From elasmobranchs to mammals the thymus arises out of two embryonic sources: (1) Its reticulum is an offshoot from some portion of the endodermal lining of the pharyngeal [illegible] dis- [illegible] in [illegible] the [illegible] blood vessels are said to have a mesenchymal origin

reticular cells of other organs. In the adult re indistinguishable from the endodermal r cells. (2) The lymphocytes (thymocytes) infiltrate the reticular network are gener- ought to migrate into it at later stages from rrounding connective tissue, although there ence that thymocytes repopulating the cortex destructive irradiation arise in the same loca-

siology. The thymus probably functions rly to the other lymphoid organs—spleen, nodes, and tonsils. Production of reserve or gamma globulin and other proteins, and fication have been suggested, but these ques- are unclarified. In addition, the thymus may ce an endocrine secretion, although intensive s to demonstrate it conclusively have thus far d. The salient fact of thymic physiology is its ssive atrophy under the influence of hormones the adrenal cortex and gonads. Other lym- c organs, particularly lymph nodes, share in atrophy but to a smaller extent.

urgical removal. In man, and most animals ed, this produces no consistent effects. Removal ng infantile stages (guinea pig) appears to re- in serious debility leading to death.

Feeding or injecting thymic extracts These hniques likewise have resulted in few consist-

ve been suggested Such experiments are com- cated by the fact that thymus is a rich food own to contain unusual quantities of proteins d vitamins B and C.

ogically secreted or injected. An "age" and an "accidental" type of atrophy can be distinguished Normally, until the age of 13, thymic weight in- creases in rough proportion to general body growth and then commences to regress under the influence of the sex hormones which appear in the blood stream at puberty (age involution). But if the pre- pubertal child is subjected to stresses which call forth excessive secretion from the adrenal cortex as in severe illness, starvation, exposure to fatigue or cold, premature thymic atrophy results. This is accidental atrophy Experimenters agree that the adrenal cortical secretion acts directly on the thy- mus. Apparently the gonads act both directly and through stimulating the adrenal cortex. The en- suing atrophy results from destruction of lympho- cytes rather than reticular cells Involution is de- layed when either the gonads or adrenals are re- moved. Adrenal removal may even produce thymic regeneration. On the other hand, injection of gon- adal hormones can induce involution in adrenalec- tomized animals, though the degree of involution is greater if the adrenals are intact.

Response to thyroid hormone. This may be essen- tially the reverse of response to gonads and adre- nal cortex. Increased thymic size accompanies thy- roid tumors, exophthalmic goiter, but this may be an indirect response through increased metabolic rate, rather than a specific reaction to thyroid se- cretion. Conversely, an inhibitory effect which the thymus appears to exert on the thyroid forms the basis of a technique which has been recently pro- posed for assaying possible thymus hormones.

Relationship between the thymus and cancer. In some strains of mice the thymus appears to play a central role in cancer produced by x-irradiation.

Relationship between the thymus and myasthe- nia gravis. This relationship is not yet clarified. This rare and often fatal disease of the muscles appears to have an association with thymic tumors. Dramatic recovery following tumor removal has been reported.

Abandoned theories Several once-prominent theories of thymic activity have failed to receive confirmation and are now questioned. These are (1) a possible relationship between the thymus and mineral metabolism, especially bone formation in mammals and shell development in birds; (2) a growth-promoting effect of thymic extracts which was described as cumulative over several genera- tions

in m

dren

SYSTEM; HEMATOPOIESIS; LYMPHATIC SYSTEM. [F.D.C.]

Bibliography: H. V. Curtis and H. Quastler (eds.), *Basic Mechanisms in Radiobiology*, Natl. Acad. Sci. Natl. Research Council, Publ. 513, 1957

Thyratron

A hot-cathode gas-filled tube with one or more grids placed between the cathode and anode to pro- vide control characteristics (Fig 1). Thyratrons operate in the gas arc region using mercury vapor where temperature control is possible, or an inert gas such as argon or xenon (*see* ELECTRICAL CON- DUCTION IN GASES, GAS TUBE) In some tubes both mercury and an inert gas are used to take advan-

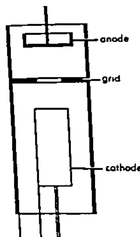


Fig. 1 Negative-grid thyratron.

tage of the desirable features of each. For applications requiring great accuracy in control, fast deionization, and very high peak currents of short duration (as in radar pulse-modulators), hydrogen gas is used.

Control action. The control action of a thyratron is quite different from that of a vacuum electron tube. In the high-vacuum tube the anode current is modulated by the action of small varying voltages applied to the grid. In the thyratron the grid serves to prevent the flow of current until its potential is made less negative than a critical value. When this occurs, electrons from the cathode are accelerated toward the anode and quickly produce an arc plasma by ionization by collision. Once the arc is established the grid has no further effect on the arc current as long as the anode is sufficiently positive relative to the cathode to supply the arc voltage When the anode potential is made zero or negative, the arc is extinguished and the grid regains control after a period of time known as the deionization time or recovery time. This is the time necessary for the ions of the residual arc plasma to be neutralized to a sufficiently low density to permit the grid to regain control. The deionization time is affected by both applied grid voltage and the grid resistance and is about 1 millisecond for most tubes, although some special tubes have deionization times as low as 10 microseconds. During the conducting period the grid acts essentially like a probe in a plasma. A negative grid in a plasma collects a positive ion current from the plasma, while a positive grid collects an electron current. By varying the grid potential relative to the cathode, a grid current–grid voltage characteristic can be plotted. The positive-ion space charge that forms about a negative probe has a thickness that can be calculated from the space charge equation using the mass of the positive ions. At quite small arc currents and with grids having small holes for the arc path, this positive-ion space-charge sheath can become sufficiently thick so as to close off the conducting path and stop the current—a mode of operation that is seldom employed. Since the function of the grid in a thyratron is to permit a relatively small negative voltage to shield the region

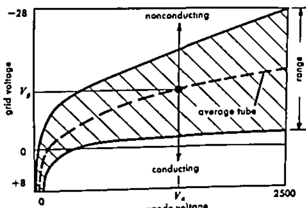


Fig. 2. Control characteristic of the negative-grid thyratron.

about the cathode from the positive field produced by the anode potential, it is evident that the grid must be made increasingly negative as the anode potential is increased.

Control characteristics. The relation between the voltage necessary to prevent conduction and the anode potential is called the control characteristic. A typical control characteristic for a negative grid thyratron (triode), such as Fig. 1, is shown in Fig. 2. The shaded area represents the range that may be expected for a group of tubes. The dotted curve represents the characteristic of an average tube. For a given value of anode voltage V_a the tube is nonconducting for grid voltages more negative than the characteristic V_g and conducting for a more positive grid voltage. The control characteristic of mercury tubes is also affected by the condensed mercury temperature, so for this type of tube the characteristics are usually given for several typical temperatures in the recommended operating range.

The construction of a shield-grid thyratron (tetrode) is shown in Fig. 3. The grid control char-

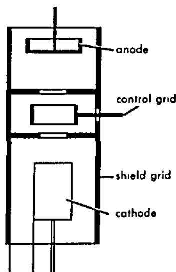


Fig 3 Shield-grid thyratron.

acteristics of this tube are similar to those shown in Fig 2, and are usually presented as a family with the shield-grid potential as a parameter. For a sufficiently negative voltage on the shield grid the tube is given positive-grid characteristics, that is, it is necessary to place a positive potential on the control grid before the tube will conduct. An important advantage of the shield-grid tube for many applications is that the control grid requires only a small current to initiate conduction. The control grid is also so placed within the shield grid that heat from the cathode and the anode cannot reach it. Thus even if active material from the cathode reaches the control grid, its low grid current property is maintained for a long period, as it never gets hot enough for thermionic emission.

Ionization time. The ionization time of a thyratron is the time required for the rated starting voltage on the grid to establish an arc in the tube The

ionization time for most thyratrons is from 0.5 to 10 microseconds, depending largely on the gas and tube construction.

Thyratron ratings. Three current ratings are given for thyratrons: (1) the average current, set by the ability of the tube to dissipate heat; (2) the peak current, the largest instantaneous current that the tube is designed to carry—the peak current averaged over a period known as the integration time must not exceed the average current; and (3) the surge current (or fault), the greatest current that may be passed by the tube under abnormal conditions such as an arc-back or a load short circuit. The life of the tube will be shortened each time this current is attained. Circuits should have sufficient impedance and be fused so that this current is never exceeded. [J D C.]

Bibliography: J. D. Cobine, *Gaseous Conductors*, 1958; W. D Cockrell (ed.), *Industrial Electronics Handbook*, 1958; K. G. Genneshauser, *Pulse Generator*, 1958.

Thyroid gland

An endocrine gland found in all vertebrates that produces, stores, and secretes the hormone thyroxin. The primary function of the thyroid is the regulation of the rate of metabolism. In humans, it is located in front of and on either side of the trachea.

The thyroid is usually a well encapsulated single

usually arising as a groove or pit at the level of the first pair of gill pouches (Fig. 1). In some fishes and amphibians the cells appear as a solid bud rather than a hollow structure. In their further development the cells of the pit or bud separate from the pharynx and migrate back to lie ultimately in a region below the ventral aorta or the trachea. During this migration they multiply rapidly and arrange themselves into elongate cords or flattened plates. These finally break up into follicles each of which gradually takes on the characteristic form of a single epithelial layer surrounding a colloid-filled lumen. In some vertebrates, of which the human is one, a contribution to the lateral portions of the thyroid is made by tissue originating from the most posterior gill pouches. It is still uncertain whether these portions (lateral thyroids, ultimobranchial bodies) carry on typical thyroid activity or are, like the parathyroid glands, distinct functional units more or less fortuitously associated with the thyroid. See ULTIMOBRANCHIAL BODIES.

Embryonic origin. However, the various vertebrate groups differ as to the time when the gland first shows evidence of secretory activity and when

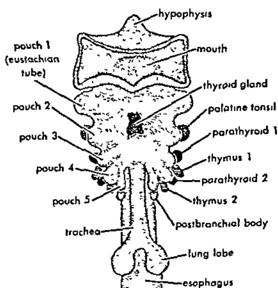


Fig. 1. Ventral view of pharyngeal region of a human embryo showing the pharyngeal pouches and their glandular derivatives; semidiagrammatic. (*From H. V. Neal and H. W. Rand, Chordate Anatomy, Blakiston, 1939*)

its hormone begins to affect development. In the frog the first organized follicles are seen in the thyroid when the larva is about 10 mm in length; while in the chick they appear after 10 days of incubation, and in the rat on the eighteenth day of development. The onset of thyroid function as judged by the ability of the gland to concentrate radioiodine precedes actual colloid appearance but there is evidence that organic combination of iodine, and therefore true hormone formation, is closely correlated with the time of colloid formation.

Thyroid and growth. The secretion of the thyroid gland plays an important part in the control of bodily growth in postembryonic stages. Growth during embryonic development seems independent of such control. Normal thyroid functioning is also essential for the successful completion of many morphogenetic processes. This is shown in anuran amphibians where the bodily changes involved in the transformation of the aquatic tadpole into the terrestrial adult frog are under thyroid control, but it is also true for those vertebrates which do not undergo metamorphosis.

Retardation of growth. The significance of the thyroid for growth is exemplified by a certain type

fested. This failure in growth, together with various abnormalities in physical and mental development, is due to congenital absence or malfunctioning of the thyroid gland. If the condition is recognized early and regular administration of thyroid hormone is instituted, normal physical growth and development can be achieved. See CRETINISM.

tage of the desirable features of each. For applications requiring great accuracy in control, fast deionization, and very high peak currents of short duration (as in radar pulse-modulators), hydrogen gas is used

Control action. The control action of a thyratron is quite different from that of a vacuum electron tube. In the high-vacuum tube the anode current is modulated by the action of small varying voltages applied to the grid. In the thyratron the grid serves *to prevent the flow of current until its potential is* made less negative than a critical value. When this occurs, electrons from the cathode are accelerated toward the anode and quickly produce an arc plasma by ionization by collision. Once the arc is established the grid has no further effect on the arc current as long as the anode is sufficiently positive relative to the cathode to supply the arc voltage. When the anode potential is made zero or negative, the arc is extinguished and the grid regains control after a period of time known as the deionization time or recovery time. This is the time necessary for the ions of the residual arc plasma to be neutralized to a sufficiently low density to permit the grid to regain control. The deionization time is affected by both applied grid voltage and the grid resistance and is about 1 millisecond for most tubes, although some special tubes have deionization times as low as 10 microseconds During the conducting period the grid acts essentially like a probe in a plasma. A negative grid in a plasma collects a positive ion current from the plasma, while a positive grid collects an electron current. By varying the grid potential relative to the cathode, a grid current–grid voltage characteristic can be plotted. The positive-ion space charge that forms *about a negative probe has a thickness that can be* calculated from the space charge equation using the mass of the positive ions. At quite small arc currents and with grids having small holes for the arc path, this positive-ion space-charge sheath can become sufficiently thick so as to close off the conducting path and stop the current—a mode of operation that is seldom employed Since the function of the grid in a thyratron is to permit a relatively small negative voltage to shield the region

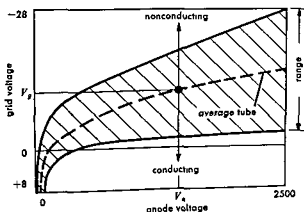


Fig. 2. Control characteristic of the negative-grid thyratron.

about the cathode from the positive field produced by the anode potential, it is evident that the grid must be made increasingly negative as the anode potential is increased.

Control characteristics. The relation between the voltage necessary to prevent conduction and the anode potential is called the control characteristic. A typical control characteristic for a negative grid thyratron (triode), such as Fig. 1, is shown in Fig. 2. The shaded area represents the range that *may be expected for a group of tubes. The dotted* curve represents the characteristic of an average tube. For a given value of anode voltage V_a the tube is nonconducting for grid voltages more negative than the characteristic V_g and conducting for a more positive grid voltage. The control characteristic of mercury tubes is also affected by the condensed *mercury temperature, so for this type of* tube the characteristics are usually given for several typical temperatures in the recommended operating range.

The construction of a shield-grid thyratron (tetrode) is shown in Fig 3. The grid control char-

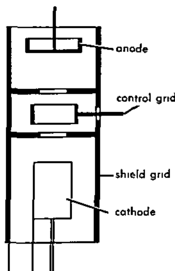


Fig. 3. Shield-grid thyratron.

acteristics of this tube are similar to those shown in Fig. 2, and are usually presented as a family *with the shield-grid potential as a parameter* For a sufficiently negative voltage on the shield grid the tube is given positive-grid characteristics, that is, it is necessary to place a positive potential on the control grid before the tube will conduct. An important advantage of the shield-grid tube for many applications is that the control grid requires only a small current to initiate conduction. The control grid is also so placed within the shield grid that heat from the cathode and the anode cannot reach it. Thus even if active material from the cathode reaches the control grid, its low grid-current property is maintained for a long period, as it never gets hot enough for thermionic emission.

Ionization time. The ionization time of a thyra*tron is the time required for the rated starting* voltage on the grid to establish an arc in the tube. The

til at metamorphosis while urodeles do not, changes in the mouthparts are very different in the two groups, and rapid hind limb growth is exclusively an anuran metamorphic feature. The precise nature of the change elicited in a tissue or organ by the thyroid hormone is not determined by any general physiological effect of the hormone but depends, rather, upon factors inherent in the tissues themselves. The characteristic differences in modes of response are genetically determined and they appear during early development as one of the features of embryonic differentiation. However, the tissues of very young larvae are not capable of manifesting any metamorphic response even when exposed to high concentrations of thyroxine Sensitivity to thyroid hormone first appears at the time when the external gills become covered by the opercular folds.

Hormonal control of thyroid Although metamorphosis is controlled by the thyroid gland, the activity of the thyroid itself is governed by a hormone of the anterior lobe of the pituitary, thyrotropin or TSH. For this reason, removal of the pituitary of a young tadpole effectively prevents metamorphosis even when the thyroid is left intact. In such an animal the thyroid develops normally but remains small and inactive. Administration of thyrotropin results in activation of the thyroid and metamorphosis follows. Thyrotropin cannot bring about metamorphosis in thyroidectomized tadpoles so it is clear that its effect is not a direct one but is exerted entirely through the thyroid. It must be assumed that the pattern of increasing activity seen in the thyroid during larval life reflects a corresponding [illegible] pro- [illegible] ges in [illegible]

Experimental methods. Surgical removal of the thyroid can be performed successfully in amphibians at any stage from the origin of the gland up to the adult organ. For most other vertebrates thyroidectomy at embryonic or fetal stages presents considerable technical difficulty Therefore most such experiments have begun with newborn mammals [illegible] be obtained, however, by the use of three indirect methods· (1) pituitary removal, (2) destruction of the thyroid by radioiodine, (3) prevention of thyroid activity by thyroid-inhibiting drugs.

Removal of pituitary In the chick embryo, removal of the pituitary can be accomplished by simply cutting away the front part of the head. This results in a high mortality but some operated chicks do survive even up to hatching The thyroids of these animals remain inactive, yet the development of the body, including some features which occur only at metamorphosis in amphibians, is not affected.

Radioiodine The use of radioiodine as a means of eliminating thyroid influence depends upon the gland's differential affinity for iodine. Radioiodine injected into the egg or administered to the fetus is concentrated in the thyroid, and if the dosage is properly adjusted, radiation damage occurs in the cells of the thyroid but not in other tissues. Obviously this method cannot be used to eliminate thyroid influence completely from the very beginning since it is not effective until the gland has become sufficiently developed to carry on its iodine-concentrating activity.

Nevertheless the results of such experiments correspond quite closely to those obtained by the use of thyroid-inhibiting drugs.

Thyroid inhibitors. Substances which inhibit thyroid function do so either by interfering with the gland's ability to concentrate iodine or by preventing the combination of iodine with protein to form the thyroid hormone. Such substances can be administered to fish or amphibian embryos by dissolving the drugs in the water in which the animals are raised. Reptilian or bird embryos may be treated by injection into the egg. For mammals, administration of the inhibitors to the mother during pregnancy is effective. Inhibition of thyroid activity by proper concentrations of thiourea, thiocyanate ion, or perchlorate ion completely prevents metamorphosis in amphibians and does not affect the growth rate. In fishes some effects upon growth have been reported and the development of secondary sex characters is prevented or delayed. In both the turtle and the chick, thyroid inhibition causes delay or complete failure of the processes of hatching and of retraction of the yolk sac into the body. Growth rate is also affected to some degree. Treatment of pregnant rats does not affect the growth or development of the young although the thyroid of the fetus shows clear evidence that its secretory activity has been inhibited.

Administration of thyroid inhibitors beginning just after birth or after hatching produces the same cretinoid effects as are seen after thyroidectomy. These include decreased growth rate, abnormal fat deposition, delayed ossification of the bones, and underdevelopment of the brain. Thus in higher vertebrates, as in the frog, many important maturation processes are carried out under thyroid control.

[W.G.L.Y.]

COMPARATIVE ANATOMY

The thyroid receives an exceptionally rich blood supply. Postganglionic fibers from the cervical ganglia and vagus enter with the blood vessels and form extensive plexuses around the smaller arteries. This innervation does not appear to be essential for normal thyroid function except as it controls the rate of blood flow through the gland. Microscopically (Fig. 3), the tissue consists of numerous vesicles lined by a single layer of epithelial cells and containing a gelatinous material called colloid This is the gland's store of hormone. The height of the secretory cells and the amount of colloid vary with the functional state of the gland.

Effect of thyroidectomy. Experimental removal of the thyroid in laboratory mammals shortly after birth produces effects corresponding to those seen in cretins. Analysis of the results of administration of various hormones to thyroidectomized animals reveals that only the thyroid hormone is effective in inducing normal maturation of skeletal system, nervous system, and other bodily structures. On the other hand, the regulation of growth processes, as distinct from morphogenetic processes, proves not to depend solely on the thyroid. Growth of thyroidectomized young rats can be markedly increased by injecting extracts of the anterior lobe of the pituitary gland. However, optimal growth rates can be attained only when a combination of thyroid hormone and anterior pituitary growth-promoting hormone is given. It is now well established that normal growth in vertebrates is under multiglandular control. It depends primarily upon balanced action of pituitary growth hormone, somatotropin, and thyroid hormone but it is also influenced by the secretions of the adrenal cortex and the pancreas. *See* HORMONE, ADENOHYPOPYSEAL.

Thyroid and morphogenesis. The role of the thyroid in morphogenesis has been most extensively studied in amphibians (Fig. 2). The metamorphosis of the tadpole into the frog involves obvious changes in body form such as loss of tail, rapid growth of limbs, protrusion of eyes, and alterations in the shape of mouth and head. Concomitant internal changes include shortening of the digestive tract, growth of lungs, degeneration of gills, and histological modifications in the tissues of such diverse organs as brain, stomach, and skin. These morphological changes are accompanied by changes in the physiological activities and capabilities of the organs concerned and thus an animal formerly adapted for life in the water now becomes adapted for terrestrial existence

In 1912 J. Gudernatsch discovered that tadpoles fed with mammalian thyroid gland undergo a precocious metamorphosis, transforming into tiny frogs at a smaller size and a younger age than do tadpoles fed any other diet. Subsequent investigators devised methods for surgical removal of the thyroid and showed that tadpoles lacking this gland fail to metamorphose although they continue to live and increase in size long after unoperated animals of the same age have transformed into young frogs. Administration of thyroid hormone to thyroidectomized tadpoles either before or after the normal time for metamorphosis elicits a prompt metamorphic response.

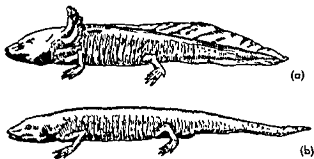
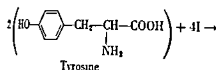


Fig. 2. The effect of iodine implantation upon the metamorphosis of the Mexican axolotl (*Amblystoma tigrinum*). (a) Untreated gilled or axolotl stage. (b) A similar specimen 30 days after the implantation of iodine crystals below the skin. Observe the atrophy of the gills and the loss of the tail fin. (*From C. D. Turner, General Endocrinology, 2d ed., Saunders, 1955*)

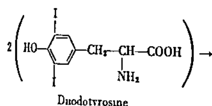
Changes during metamorphosis. The transformations that take place at metamorphosis have been studied in detail. For many of them it is known not only how they occur during normal development but also how they are influenced by thyroid removal, generalized thyroid administration, and localized application of thyroid substance by implantation of hormone-containing pellets. Most metamorphic events prove to be under direct thyroid control. A few, however, such as the perforation of the operculum to permit emergence of the forelimb, are at least partially dependent upon other factors. Since these other factors are themselves influenced by the thyroid, this is an example of indirect thyroid control. It is of interest that the development of the gonads is not dependent upon the thyroid. Precocious metamorphosis induced by early thyroid administration is not accompanied by precocious sexual maturity, and the ovaries and testes of thyroidectomized tadpoles develop normally.

The bodily alterations that characterize metamorphosis do not occur simultaneously. Some changes, like growth of the hind limbs, begin early and require several weeks or months for their completion; others, such as tail resorption and loss of tadpole mouthparts, occur later and more rapidly. The sequence and spacing of metamorphic events is determined by two factors: (1) the changing rate of secretory activity in the thyroid and (2) differences in the rates of response of the tissues affected. The thyroid of the tadpole maintains a low level of hormone production during early larval life Secretory activity begins to increase gradually just before the hind limbs start to enlarge and reaches a peak near the time of metamorphic climax. The various tissues and organs which undergo changes at metamorphosis differ in their rates of response to a given concentration of thyroid hormone Thus, during the early phases of metamorphosis, when the amount of hormone released into the blood stream is low, only a few changes, notably growth of the hind limbs, are seen. Later, as the concentration of hormone in the blood increases, other changes appear in a definite sequence and proceed with a characteristic speed.

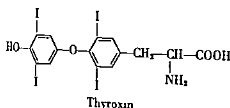
Tissue response The responses of the tissues to the metamorphosing agent are quite diverse. Some organs grow, others are resorbed, still others show specific histological alterations. Moreover the response of an individual organ is often not the same in different groups of amphibians. Anurans lose the



Tyrosine



Diiodotyrosine



Thyroxin

Thyroxin is the principal hormonal constituent of the blood; the other iodothyronines being present only in minute amounts. Once thyroxin enters the circulation, it becomes promptly bound to plasma protein (probably an α-globulin) The thyroid hormones are metabolized principally in the liver and kidneys In the liver thyroxin and to a lesser extent triiodothyronine are conjugated as glucuronides, and then freed into the intestine via the bile; both compounds may undergo oxidative deamination in the liver to form the corresponding pyruvic acid derivatives. In the rat, much of the hormonal iodine is excreted in the feces as thyroxin, but in man it is excreted chiefly by the kidney as iodide.

The normal thyroid has a remarkable capacity to take up inorganic iodine from the blood and fix it in organic forms. The enzymes of the gland which affect these transformations are unknown While the thyroid hormones are known to increase the quantities of many important tissue enzymes, and to exert generalized actions within organisms, research has failed to reveal what might be a specific thyroid effect upon a target tissue or organ It is certain that thyroid function [illegible] ly integrated [illegible] e body. [illegible] n elucidations, and [illegible] known. Deficiency may be produced by surgically removing the gland or by administration of antithyroid substances such as thiourea, thiouracil, and sulfa drugs These compounds apparently exert their action by preventing the normal synthesis of hormones by the gland In human beings, thyroid tissue may be congenitally absent or subnormal in amount, the study of such cases is instructive

Hypothyroidism. Hypothyroidism is reflected by reduced metabolic rate in warm-blooded animals. There appears to be a gradual reduction in certain cellular enzymes and, as a consequence, all metabolic processes are retarded. This alteration in the rate of tissue oxidation appears to be the most specific and basic role of the thyroid hormones. Blood pressure, cardiac output, heart rate, and circulation time are all reduced. The motility of the gastrointestinal tract and the secretion of digestive juices are diminished. Body temperature and rate of breathing are typically below normal. Hypothyroidism in cold-blooded animals appears not to lower the basal metabolic rate (BMR), though it delays somatic differentiation in these species. Thyroid compounds accelerate metamorphosis in Amphibia and this effect is widely used for bioassay of these hormones

Deficiency in young individuals impairs growth and results in a type of dwarfism. Cretinism is due to thyroid deficiency during infancy and childhood, and is characterized by stunted growth, mental impairment [illegible] adult man (myxedema) is accompanied by thickening and puffiness of the skin and subcutaneous tissues giving the body an edematous appearance, scaliness of the skin and loss of hair, and often extreme somnolence.

In addition to mental retardation which may approach imbecility or idiocy, the diminished frequency of the α-waves of the brain indicates subnormal functioning of the nervous system. Sluggish tendon reflexes suggest defects in the neuromuscular apparatus.

Alterations in intermediate metabolism are not very conspicuous. However, the rate of absorption of fatty acids and glucose from the alimentary tract is below normal. Blood cholesterol is high, and fasting nitrogen excretion diminishes as the BMR falls

Hyperthyroidism Hyperthyroidism may result from toxic goiter or from excessive administration of the hormone In general the effects are the reverse of those listed above. The clinical manifestations include nervousness or other psychic disturbances, increased sweating, mild to extreme loss of weight, inability to sleep normally, muscular weakness, varying degrees of diarrhea, tremor of the hands, high BMR, a rapid heart rate, and high titers of protein-bound iodine in the plasma. Protrusion of the eyeballs (exophthalmus) occurs in some hyperthyroid patients, but this defect is thought to be due to the thyrotrophic hormone of the pituitary gland rather than to a direct effect of the thyroid hormones Subtotal thyroidectomy, iodine therapy, or treatment with antithyroid drugs all serve to relieve the hyperthyroid state.

Goiter. This is a visible enlargement of the thyroid gland, and occurs frequently in man and other vertebrates. Enlargement of the gland does

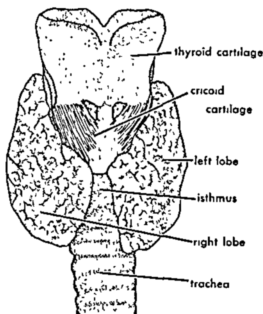


Fig 3. Ventral view of human thyroid gland shown in relation to trachea and larynx. (*From C. K. Weichert, Anatomy of the Chordates, 2d ed, McGraw-Hill, 1958*)

Thyroid secretion is controlled largely by the thyrotropic hormone from the anterior pituitary.

Fishes. In most fishes the thyroid is represented by diffuse masses of follicles scattered along the large arteries which enter the gills ventrally, or tucked between muscles in the pharyngeal floor. In elasmobranchs the follicles are aggregated in a single mass. In the remaining fishes (except lampreys) and in higher vertebrates a pair of thyroid masses, often connected by a median strand (isthmus), usually occurs. Minute accessory thyroid masses are common.

Amphibians. In amphibians the thyroid lobes lie under cover of certain muscles of the buccal or pharyngeal floor near the caudal angle of the jaws

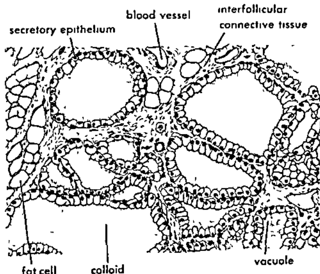


Fig. 4 Histologic features of the normal thyroid gland of the rat. (*From C. D. Turner, General Endocrinology, 2d ed., Saunders, 1955*)

(anurans), or at the base of the branchial arches (urodeles).

Amniotes. In amniotes the two oval or elongated glands lie against the trachea, immediately below the larynx in mammals (Fig. 4), part way down the trachea in lizards, still farther caudad in other reptiles, or just above the bifurcation of the bronchi in birds. In most mature reptiles the gland is unpaired. The thyroid glands of man are attached to the thyroid (shield-shaped) cartilage of the larynx.

Phylogeny. A clue to the phylogenetic history of the thyroid is found in lampreys. In marine species an elongated rod of cells capable of selectively absorbing iodine-rich substances is located in the pharyngeal floor between the second and fifth gill pouches In larval brook lampreys the iodine-capturing cells are part of a complicated subpharyngeal gland (endostyle) which evaginates from the embryonic pharyngeal floor At metamorphosis, the gland loses its connection with the pharynx and remains as isolated thyroid masses underneath the pharynx. Embryonic origin of the thyroid of higher vertebrates as a pharyngeal outpocketing probably represents a recapitulation of the phylogeny of the gland. Although the embryonic pharyngeal connection is usually lost, a duct remains patent in some elasmobranchs. Even in man remnants of the duct may persist. [C.K.]

PHYSIOLOGY

The thyroid, a gland of internal secretion, manufactures, stores and secretes a hormone that regulates the metabolic rate. The lobes are composed of follicles lined with secretory cells and filled with a substance called colloid, the storage product of the secretory cells.

Thyroid metabolism. Thyroid metabolism has been studied through the techniques of chromatography and autoradiography of compounds labeled with the radioactive isotope of iodine, I^{131}, and a variety of iodinated compounds have been identified in the thyroid, blood, and tissues. With the discovery of triiodothyronine, which has greater biological potency than thyroxin, a question arises concerning the nature of the thyroid hormones. The final answer is not yet available. but it is certain that thyroid actions are mediated by multiple hormonal agents.

Amino acids are essential for the biosynthesis of hormones of the pituitary gland, pancreas, adrenal medulla, and thyroid The biosynthesis of thyroxin makes essential the provision of tyrosine (or phenylalanine) from which the hormone is probably produced. The formation of thyroxin from tyrosine and iodine may be represented as shown in the accompanying equations.

Thyroxin can be prepared in the laboratory by treating tyrosine-containing proteins, for instance casein, with elemental iodine. This process apparently is not enzymatic Such iodinated thyroproteins may be prepared cheaply and used for treatment of farm animals.

before any influence on the pulse is detectable. Just as is the case with many of the endocrine glands that work in pairs, a balance in the functioning of the normal organism is maintained by the interaction between these two glands, the thyroid and the pituitary. One of the most common of the disorders of the thyroid, simple goiter, or enlarged thyroid gland, is brought about by the working of this reciprocal mechanism. If there is in the system a deficiency of iodine with which to synthesize the thyroid hormone, often due to a deficiency of this

Under this stimulation, the area of secretory surface of the thyroid gland enlarges, so that the gland can attempt to keep up with the demand for hormone. Often the output of hormone from the enlarged gland is normal, and if the temporary deficiency is corrected, the gland itself returns to normal size. If, despite the increased pituitary stimulation, the thyroid is unable to compensate satisfactorily, the gland, unable to stand the continued physiologic strain, finally becomes exhausted, and irreparable atrophy of its cells results. *See* ATROPHY.

With exophthalmic goiter or Graves' disease, there is not only extensive enlargement (hypertrophy) of thyroid tissue with abnormal multiplication of the number of cells (hyperplasia), but also excessive production of thyroid hormone. Sometimes masses of tumorous tissue are found embedded in the gland. *See* HYPERPLASIA, HYPER-
...HY. [C.H.L.]

...liography: A. Gorbman, Some aspects of the ...parative biochemistry of iodine utilization and ...olution of thyroidal function, *Physiol. Revs.*, ...36–346, 1955; J. Roche and R. Michel, Nature ... metabolism of thyroid hormones, *Recent ...gr. in Hormone Research*, 12:1–26, 1956; C. D. ...ner, *General Endocrinology*, 2d ed., 1955; R. L. ...lterson, *Endocrines in Development*, 1959

...yroid gland disorders

...e most common thyroid disorders are neoplasias ...d those of dysfunction produced by inflammation associated with tumors.

Simple goiter is an enlargement of the gland ...om unknown causes and not from frank iodine ...ficiency. It is seen most often in adolescent fe...ales, thereby suggesting a hormonal imbalance; ...other cases diet or hereditary factors have been ...mplicated. There is a gradual enlargement of the ...hyroid, usually without other symptoms. The thy...oid may later decrease in size or change to a nodu...ar goiter in later life. There is no dysfunction of ...the gland in most cases. *See* HORMONE; THYROID GLAND.

Endemic goiter is enlargement of the thyroid resulting from iodine deficiency of food and water. It is found in more than 10% of a local population. In certain areas, some Alpine villages, for example, most of the population is affected. No glandular dysfunction is found unless severe deficiency is present. Such dysfunction is termed cretinism, a form of hypothyroidism. Various forms of iodine, added to salt or candy, are used both as effective prevention and in mass treatment.

Hyperthyroidism, or excess secretion of thyroid hormone, is also known as thyrotoxicosis, Graves' ... exophthalmic goiter. Al-

symptoms, some of which are quite characteristic. The most common are fatigue, muscle weakness or tremor, rapidity or irregularity of the heartbeat, and heat intolerance. Despite an increase in appetite, there is usually a weight loss. In classical Graves' ...
forward, ...
Other symptoms ... cur form the basis for further classification of this disorder. The basal metabolic rate is almost always elevated as is the absorption of radioactive iodine. These two diagnostic tests are used in conjunction with the clinical history to establish the diagnosis.

Inflammation of the thyroid, or thyroiditis, oc-

if the disease becomes chronic and excessive inflammatory tissue is formed.

Thyroid neoplasia includes benign nodules and carcinomas, as well as less common benign and malignant growths. By far the most frequently seen is nodular goiter, consisting of small masses of new glandular or supporting tissue. There are several histologic types of carcinoma of the thyroid, each with its own characteristics and prognosis. The variability of histologic pattern often complicates diagnosis. *See* ONCOLOGY.

Hypothyroidism includes any condition in which insufficient thyroid hormone is produced or circulated. Cretinism results from inadequate iodine supply in fetal or infant life. There is retardation of mental and skeletal development which, in some cases, may be partially alleviated by thyroid therapy. Other forms result from chronic inflammation, hereditary defect, or enzymatic and nutritional derangement.

Myxedema, of either juvenile or adult type, may follow an inflammation, or other specific event, but often occurs without apparent cause. The patient has a dry, coarse, and roughened skin and a puffy, swollen face. The hair is also dry and coarse and the tongue may be characteristically enlarged and somewhat protruding. There is a tendency toward obesity. Constipation is frequently a major problem. The heart, blood pressure, and reflexes may be affected. There is no actual mental loss but sluggishness and apathy may be well marked.

Other thyroid disorders include congenital defects and certain regressive changes like atrophy which may follow a decrease in pituitary stimulation or occur as a concomitant of advancing age. [E.C.ST.]

not reflect the amount of thyroid hormones being released; hormonal output may be subnormal, normal, or above normal. Toxic goiters, whether diffuse (Graves' disease) or nodular (toxic adenoma), liberate excessive hormones; nontoxic goiters secrete normal or subnormal amounts.

Simple goiter is the most common type and is associated with a relative or absolute lack of iodine in the food or drinking water. It is known to have been endemic for centuries in certain areas (goiter belts) where the soil is deficient in iodine. The prophylactic use of iodine, together with improved transportation of foods, has greatly reduced the incidence of this disease.

If the gland is not dangerously enlarged, and if the BMR is within normal limits, this type of goiter may not threaten an individual's health. However, many clinicians feel that untreated simple goiters are likely to become overtly hypofunctional. Hypothyroid diseases (cretinism and myxedema) are known to be more frequent in the goiter belts than elsewhere.

Since the functional state of the thyroid is conditioned by the thyrotrophic hormone, it is probable that some goiters result from pituitary defects rather than thyroid defects.

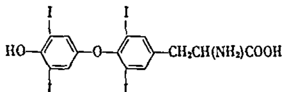
Goitrogenic agents. Goitrogenic (antithyroid) agents interfere in some manner with the synthesis of thyroid secretions, the resulting hypothyroid state causes the release of excessive pituitary thyrotropin which produces thyroid enlargement (goiter). The fact that goitrogens do not produce thyroid enlargement in hypophysectomized animals shows that the goiter is mediated by the pituitary. Certain plant foods, notably the cabbage and turnip families, brassica seeds, and soya beans contain goiter-producing substances. The sulfonamides, substituted thioureas, thiouracil and its derivatives are important in both experimental work and in clinical medicine. Such substances have been employed with considerable success in the treatment of hyperthyroidism in human patients. [C.D.T.]

BIOCHEMISTRY

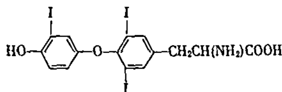
The size of the normal thyroid is subject to more variation than other organs in the body, fluctuating with age, reproductive state, diet, and external environment. The average weight in the adult human subject is from 25 to 40 g.

Iodine is an essential ingredient that must be present in the body for the manufacture of the thyroid hormone. Although this element is present in all mammalian tissue, the thyroid has the greatest ability to trap iodine, binding one-third to one-quarter of the total amount of this element in the body. The gland takes up iodine and fixes it extremely rapidly. As the first step in the synthesis of the thyroid hormone, iodine is extracted from the circulation, and in the gland it combines with the amino acid tyrosine. The colloid that fills the follicles of the gland is a protein globulin, known as thyroglobulin; this is the storage form of the hormone. It cannot pass through the membrane of the cell into the circulation in this form, however, because of the large size of the molecule, so the thyroglobulin is broken down by an enzyme system into its constituents, one of which is thyroxin. Thyroxin is the form in which the hormone enters the circulation. The thyroxin may be degraded or changed by the tissues to still another compound, L-triiodothyronine, which is biologically more active than thyroxin itself.

By enzymic digestion of thyroglobulin, E. C. Kendall in 1914 isolated the active component, thyroxin, in pure crystalline form. In 1926, C. R. Harrington established by both degradation and synthesis the structural formula of thyroxin as 3,5,3',5'-tetraiodo-L-thyronine:



In 1951–1952, R. Pitt-Rivers, J. Roche and their coworkers demonstrated the occurrence of another iodine-containing compound in the thyroid extract, which they identified as 3,5,3'-triiodo-L-thyronine:



The biological activity of L-triiodothyronine is about three times that of L-thyroxin.

The particular metabolic effect of the thyroid hormone is regulation of the rate of energy exchange. The customary test for the functioning of the thyroid is measurement of the basal metabolic rate, that is, the quantity of heat liberated by an organism at relative rest. There are many corollary effects that follow from this primary metabolic function of the thyroid. The metabolism of carbohydrates, proteins, and fat is influenced by the gland, for example, too much thyroid in the circulation (hyperthyroidism) intensifies the symptoms of diabetes mellitus (*see* INSULIN; PANCREAS). The thyroid also exercises an effect on growth and sexual differentiation. In addition, the lowered or heightened metabolic rates affect in turn the functioning of the heart and the circulatory mechanism.

The thyroid gland is under the regulation or supervision of the pituitary gland. Thyrotropin or thyroid-stimulating hormone from the anterior lobe of the pituitary stimulates the thyroid, which is its target organ, to manufacture and secrete its hormone. It is interesting that the injection of pituitary thyrotropin causes an immediately observable rise in the pulse rate, whereas when the thyroid hormone thyroxin is injected, about 6 hours elapse

. . of considerable importance.

ceo!
tan
Mac
nps
or p......
oung are reared.

e family Thripidae is the largest, most impor-, and most injurious family of thrips This fam-has about 33 genera with 200 species They feed n leaves, fruits, buds, and flowers and often se serious crop losses. Some species have an ex-dingly wide host range and may feed and breed more than 100 kinds of plants. *See* INSECTA; [E O.E.]
ECTICIDE.

Bibliography: J R. Watson, Synopsis and cata-gue of the Thysanoptera of North America, *Univ. lorida Agr. Expt. Stas. Bull.*, 168, 1923.

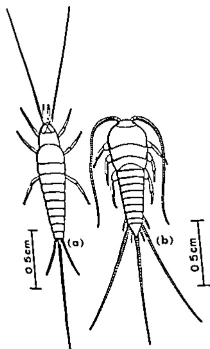
hysanura

An order of small, primitive, apterous insects, hav-ng flattened, naked, or scaly bodies, a soft integu-ment, and primitive metamorphosis The mouth-parts are mandibulate The maxillae are long, as are the many-segmented antennae. Compound eyes may be well developed, vestigial, or absent and ocelli may be present or absent. The coxae are small and the tarsi are two- or three-segmented, with two or three claws. Abdominal styliform ap-p...

most primitive of all insects Their distribution is world-wide. They are white, gray, brown, or other-wise pigmented to harmonize with their immediate surroundings on the ground, among dry or wet leaves, rocks, and vegetation. They now number ap-proximately 3 subfamilies, 30 genera, and 150 spe-cies. They are most abundant in the Palearctic re-gions, where nearly half the known species have been collected.

... portant species are well-known

tan in distribution, being known in ... Europe, China, Japan, and the Hawaiian Islands. Another species is the fire brat, *Thermobia domes-tica* (Packard), which now commonly occurs in the habitations of man throughout the world. They are also found in the nests of termites, especially in South America. *See* APTERYGOTA; INSECTA. [E.O E.]

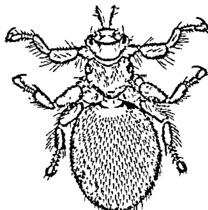


Examples of Thysanura (a) *Machilis maritima* Leach after Lubbock) (b) *Acrotelsa collaris* Escherich (after Escherich, from E. O Essig, College Entomology, Mac-millan, 1942)

Tick

Any member belonging to the suborder Ixodes, or-der Acarina, class Arachnida, phylum Arthropoda.

Ticks are of considerable importance because they suck the blood of their hosts When large num-



The sheep tick, *Melophagus ovinus;* length about ¼ in. (From E L Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

bers of ticks attack the same host they may cause serious weakening They are also the vectors of a number of diseases of man and other mammals, the best known of which are Rocky Mountain spotted fever of man, Texas fever of cattle, and tularemia, which attacks a wide variety of animals including man

Structure. Although they are separated from the various suborders of mites on the basis of fine tech-nical differences, ticks are generally much larger than the mites, and possess a leathery skin, which is lacking on the mites There are two families of ticks, the Ixodidae, or hard ticks, and the Argas-idae, or soft ticks. The most obvious difference is the presence of a scutum, or dorsal shield, on the hard ticks, which is lacking in the other family.

Reproduction. All ticks have a similar life his-tory, but with some variations in details. There is

Thysanoptera

This order consists of mostly small, slender, cylindrical, terrestrial insects that range from 0.6 mm to 14 mm in length, commonly called thrips. They have incomplete metamorphosis, although in a few species a complex type is approached. Thrips live on plants and rank high among the destructive pests of crops. The order contains upwards of 3400 species and dates back to the Jurassic.

Thrips are generally found upon all types of vegetation and are perhaps most abundant in flowers and on the leaves of host plants, although they feed on twigs and may seriously scar or deform fruit. They have a tendency to be gregarious and the effect of their feeding is the destruction of the plant epidermal cells, resulting in a silvering of the leaves, fruit, and stems. The attacked area is covered with tiny spots of black excrement which give the injured surface a speckled appearance. Often the attacked parts become deformed and dry, and they may drop. The flower-feeding forms frequently destroy the buds and blossoms completely. Some deformation of fruiting bodies is caused by punctures made in developing fruits by the ovipositor of egg-laying females. Many species that normally feed on grasses and weeds in uncultivated areas migrate to cultivated fields and orchards when the native vegetation begins to dry up. These individuals frequently do a great deal of damage, particularly to the developing fruit in orchards. Some species of thrips are important vectors of destructive virus diseases of cultivated crops.

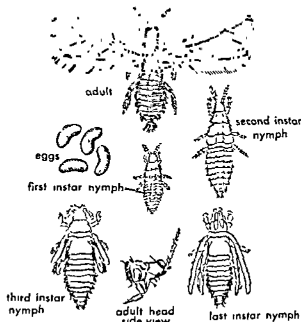
Economically important thrips

Common name	Scientific name	Plant host
Bean thrips	*Hercothrips fasciatus* (Pergande)	Legumes
Greenhouse thrips	*Heliothrips haemorrhoidalis* (Bouché)	Many plants
Citrus thrips	*Scirtothrips citri* (Moulton)	Fruit and foliage
Wheat thrips	*Frankliniella tritici* (Fitch)	Grains and flowers
Onion thrips	*Thrips tabaci* Lind	Beans onion, tobacco cabbage, and tomatoes
Pear thrips	*Taeniothrips inconsequens* (Uzel)	Fruit and foliage
Olive thrips	*Liothrips oleae* (Costa)	Olive
Christmasberry thrips	*Liothrips ilex* (Moulton)	Christmasberry
Lily thrips	*Liothrips vaneecki* Priesner	Bulbs of lily
Corn thrips	*Limothrips cerealium* Holiday	Grains
Striped thrips	*Aeolothrips fasciatus* (L.)	Trees shrubs, also predaceous on other thrips and aphids
Gladiolus thrips	*Taeniothrips simplex* Morison	Plants, flowers, and corms
Six-spotted thrips	*Scolothrips sexmaculatus* (Pergande)	Predaceous on plant mites

The mouthparts, which are situated far back on the underside of the head, are cone-shaped and modified for piercing, chafing, and sucking. The antennae are 6- to 9-segmented. Wings are rarely missing or abortive and usually there are two pairs, which are narrow, have few or no veins, and are fringed with many long or short hairs or bristles. At rest, the wings are folded flat over the back with the hairs or bristles lying against the wing margins. The legs are short and terminate in 1- or 2-segmented tarsi, the latter being more common. Further, the tarsi have one or two claws and end in an inflatable membranous bladder which is characteristic of the order. The color range is restricted to various shades of yellow, tan, orange, reddish brown, or wholly black, or to combinations of these. The wings may be the color of the body, paler and streaked, or mottled with other darker shades The adults crawl or run with a slow even gait or run very rapidly and then leap to fly away with great agility. Some species, in crawling, curve the abdomen upward, a pose assumed by the young as well as the adults.

Parthenogenesis is common in the order, and in some species, notably the greenhouse thrips, *Heliothrips haemorrhoidalis* (Bouché), and the pear thrips, *Taeniothrips inconsequens* (Uzel), males are either exceedingly rare or unknown. Depending upon the species, there may be 1–10 or more generations each year. There are two suborders, the Terebrantia and the Tubulifera. Species belonging to the Terebrantia have the ovipositor modified into a sawlike organ for inserting the tiny eggs into plant tissue, while the Tubulifera lack such an ovipositor and lay their eggs on the surface. Members of the genera *Actinothrips, Diceratothrips, Elaphrothrips*, and *Zeuglothrips* are viviparous, that is, give birth to living young.

A few species are beneficial since they are predaceous and feed upon pest populations. The six-spotted thrips, *Scolothrips sexmaculatus* (Pergande), is a valuable and abundant species that preys on a number of species of spider mites. The black hunter, *Leptothrips mali* (Fitch), is another



Pear thrips, *Taeniothrips inconsequens* (Uzel), life cycle. (From C. L. Metcalf and W. F. Flint, *Destructive and Useful Insects*, 3d ed., McGraw-Hill, 1951)

some European countries and is caused by appropriate species of indigenous ticks. In Australia, *Ixodes holocyclus* causes frequent cases in dogs, and occasionally in man, and paralysis has been known to progress even after removal of ticks; serum from recovered dogs has been shown to have some curative properties and was at one time produced for treatment. *Ixodes pilosus* is associated with the disease in South Africa.

Since 1903, over a hundred human cases and many outbreaks in cattle, sheep, and even domesticated bison, have been recorded in the northwestern states and southern British Columbia, due to attacks by *Dermacentor andersoni*. Several human

prevalence. Incidence is highest in children from 1-5 years of age, with more than twice as many girls affected as boys, presumably because their longer hair conceals feeding ticks However, the sex ratio is reversed among the fewer cases in adults due to difference in exposure. Young to yearling stock are most prone to the disease in sporadic years for reasons still unknown.

The related American dog tick, *D variabilis*, has paralyzed persons and dogs in the southeastern United States, and a few cases have been associated with the lone star tick, *Amblyomma americanum*

The female tick requires 4-5 days of feeding before initial symptoms appear and the disease progresses rapidly in the next 2-4 days. Experimental, fatal paralysis has recently been produced by *D andersoni* females on woodchucks, ground squirrels, wood rats, hamsters, guinea pigs, dogs, and lambs. Signs of the disease occur within a few hours of transfer of partially fed females, but not males, to fresh animals. The toxic principle has so far not been isolated. [C.B P.]

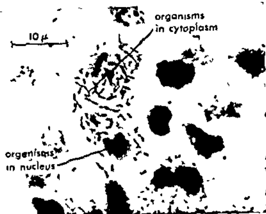
Bibliography: K. H. Abbot, Tick paralysis, a review, *Proc. Staff Meetings Mayo Clinic*, 18:39-45, Feb. 10, 1942, 18.59-64, Feb 24, 1942, L. E Hughes and C. B. Philip, Experimental tick paralysis in laboratory animals and native Montana rodents, *Proc Soc. Exptl Biol Med*, 99(2):316-319, 1958.

Tick typhus, North Queensland

carried by the tick, *Ixodes holocyclus* The symptomatology, including eschar, is similar to rickettsialpox, but differentiation is based chiefly on epidemiologic, serologic, and (in laboratory animals) immunologic grounds. The Weil-Felix OX_{19} is positive. See RICKETTSIOSES; SEROLOGY [C B P]

Tick typhus, Siberian

A relatively benign, rash- and eschar-producing, spotted feverlike disease in northern Asia, caused by bacterialike microorganisms, *Rickettsia siberica*. The disease is transmitted by four species of



Rickettsia siberica, causative agent of Siberian tick typhus from scrotal sac of infected guinea pig. (Photomicrograph by P. F. Zdrodovskiy)

Dermacentor and two of *Haemaphysalis*. See ACARINA.

Clinically, Siberian tick typhus resembles most closely fièvre boutonneuse, including low mortality, but epidemiologically, it is like American spotted fever, with the primary natural cycle between ground squirrels and other small rodents, and their immature tick parasites (see SPOTTED FEVER, ROCKY MOUNTAIN). Cross-protection tests with experimental vaccines also suggest a closer relationship to *R. rickettsii* than to *R. conorii*. Weil-Felix OX_{19} and complement-fixation tests are diagnostic The bulk of cases occur in May and June. Foci of infection are in grassy meadowlands and steppes, or brushy hillsides. See FIÈVRE BOUTONNEUSE; RICKETTSIALES; RICKETTSIOSES [C.B.P.]

Tickle

A lively pattern of cutaneous sensation involving rapid moment-to-moment intensitive variations of a light-touch or contact quality The reflex arousal of withdrawal responses adds a kinesthetic component to the total feeling pattern (see KINESTHETIC SENSATION)

On purely observational grounds, tickle seems to be most closely allied to pressure sensitivity. The most commonly observed circumstance for its evocation is multiple stimulation with the lightest contactors (a wisp of cotton or a feather tip), in a region rich in touch receptors, such as the lips (see TOUCH). The current conception of tickle, therefore, is that it is a complex spatiotemporal pressure pattern, resembling somewhat vibration (a more regular but less lively whirring) and formication (so named because it recalls the crawling of ants over the skin).

There is an appreciable body of evidence, both experimental and clinical, that also links tickle with pain (see PAIN, CUTANEOUS). If one induces pressure anesthesia of the hand by prolonged application to the upper arm of a blood pressure cuff, pain sensitivity still being retained, it is possible to evoke tickle by stroking the forefinger with a stiff nylon thread. This feeling readily passes

always an egg; a six-legged larva, commonly called the seed tick; the nymph, or yearling tick, which is eight-legged; and the adult, which also has eight legs. Eggs are usually laid on the ground. The larva reaches its host in a variety of ways, but after feeding on the host for a time, it drops to the ground where it molts and becomes the nymph. The nymph finds a new host, feeds for a variable time, and drops to the ground again. The nymph molts into the adult, which can always be distinguished by the genital pore, lacking on the nymph. The adult attaches to a third host on which the tick feeds until it attains sexual maturity and mates. After mating the female drops to the ground, deposits its eggs, and dies. *See* ARACHNIDA; ARTHROPODA; MITE; PARASITOLOGY. [J D B.]

Tick bite fever, South African

An infectious tick-borne disease of man which is similar to fièvre boutonneuse but has some biological and ecological differences. A similar, or the same, disease occurs in East Africa where it is known as Kenya tick typhus. These have been considered to be variants of fièvre boutonneuse with different epidemiology, but laboratory studies raise suspicion that the agents are at least a different subspecies or even a separate species. If the agent is a separate species, then the name, *Rickettsia pijperi* do Amaral and Monteiro, would apply to it.

The chief vector in Kenya is the dog tick, *Haemaphysalis leachi*, and in South Africa, this vector plus *Rhipicephalus evertsi*. In both areas, the disease may be acquired in domiciliary premises or from rural and bushveld sources. *See* RICKETTSIALES; RICKETTSIOSES. [C B P.]

Tick fever, Colorado

A mild, febrile viral disease transmitted by wood ticks. The virus is antigenically distinct from other arthropod-borne viruses. It is pathogenic for hamsters and mice, and multiplies in chick embryos. *See* CULTURE, EMBRYONATED EGG.

After 4–6 days' incubation, sudden onset comes with chills, fever, severe aches, sometimes nausea, symptoms and fever are often diphasic. Occasional cases are reported with more severe symptoms involving the central nervous system. *See* CENTRAL NERVOUS SYSTEM.

Diagnosis is made by complement-fixing or neutralizing antibody rises in serum. Immunity is lifelong. Ticks are a true reservoir, because the virus is transmitted to offspring transovarially. *See* COMPLEMENT-FIXATION TEST; NEUTRALIZING ANTIBODY.

Control is by avoidance of, or protection against, tick bites. *See* ANIMAL VIRUS; TICK. [J.L.M.]

Tick paralysis

A loss of muscle function or sensation in man or certain animals following the prolonged feeding of female ticks (*see* IXODIDES). Paralysis, of Landry's type, usually begins in the legs and spreads upwards to involve the arms and other parts of the body. Evidence suggests that paralysis is due to a neurotoxin formed by the feeding ticks rather than the result of infection with microorganisms. *See* TOXIN, BACTERIAL.

Present information indicates that (1) in endemic areas, resistance to tick paralysis is found in older animals as well as certain animal species, (2) some animals which have recovered are not immune, (3) *only occasional female, but not male,* ticks may induce the disease under favorable host or environmental conditions, the specific requirements being as yet unknown, (4) paralysis of adult persons and domestic animals as large as a 1000-lb bull has resulted from only one partially engorged tick, usually but not necessarily attached about the head or upper spine, (5) death may ensue if respiratory centers are reached by the ascending paralysis before the offending tick completes feeding or is removed, and (6) in most areas, recovery is prompt, a matter of hours, when the ticks are removed. It is highly important, therefore, that the disease be properly and promptly diagnosed and search for the tick be instituted immediately.

The disease has been reported in North America, Australia, South Africa, and occasionally in

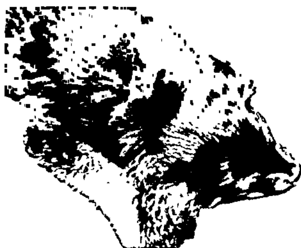


Fig. 1 Tick paralysis in western Montana puppy caused by *Dermacentor andersoni* female tick, partially engorged near shoulders. (*Rocky Mountain Laboratory*)

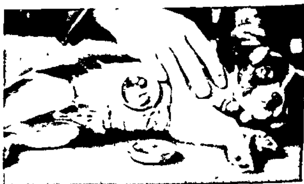


Fig. 2. Experimental tick paralysis in puppy due to two partially, and one nearly fully, engorged female ticks in capsule bandage on belly. (*Rocky Mountain Laboratory*)

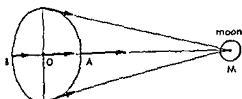


Fig. 1. Schematic diagram of the lunar gravitational force on different points in the earth.

rection and magnitude of the lunar attractive force, from one point of the earth's mass to another, which gives rise to the tidal stresses.

According to Newton's laws, the moon attracts every particle of the earth with a force directed toward the center of the moon, with magnitude proportional to the inverse square of the distance between the moon's center and the particle. At point A (see Fig 1) the moon is in the zenith and at point B the moon is at nadir. It is evident that the upward force of the moon's attraction at A is greater than the downward force at B because of its closer proximity to the moon. Such differential forces are responsible for stresses in all parts of the earth. The moon's gravitational pull on the earth can be expressed as the vector sum of a constant force, equal to the moon's attraction on the earth's center, and a small deviation which varies from point to point in the earth (Fig 2). This small deviation is referred to as the tide-generating force. The larger constant force is balanced completely by acceleration (centrifugal force) of the earth in its orbital motion around the center of mass of the earth moon system, and plays no part in tidal phenomena. See GRAVITATION

The tide-generating force is proportional to the

important, insofar as tides are concerned, as the sun, although the latter's direct gravitational pull on the earth, which is governed by an inverse-square law, is about 180 times the moon's pull

The tide generating force, as illustrated in Fig. 2, can be expressed as the gradient of the tide-generating potential, which has the form

$$\psi = \frac{3}{2}\frac{\gamma M r^2}{c^3}\left(\frac{1}{3} - \cos^2\lambda\right) \qquad (1)$$

the small number r/c have been neglected As ψ depends only on the space variables r and λ, it is symmetrical about the earth moon axis

It helps one visualize the form of the tide-generating potential to consider how a hypothetical "inertialess" ocean covering the whole earth would respond to the tidal forces. In order to be in equilibrium with the tidal forces, the surface must assume the shape of an equipotential surface as determined by both the earth's own gravity and the tide-generating force. The elevation of the surface is given approximately by

$$\bar{\zeta} = -\frac{\psi}{g} + \text{const} \qquad (2)$$

where ψ is evaluated at the earth's surface and g is the acceleration of the earth's gravity. The elevation $\bar{\zeta}$ of this hypothetical ocean is known as the equilibrium tide. Knowledge of the equilibrium tide over the entire earth determines completely the tide-generating potential (and hence the tidal forces) at all points within the earth as well as on its surface. Therefore, when the equilibrium tide is mentioned, it shall be understood that reference to the tide-generating force is also being made.

Harmonic development of the tide. The equilibrium tide as determined from relations (1) and (2) has the form of a prolate spheroid (football-shaped) whose major axis coincides with the earth-moon axis The earth rotates relative to this equilibrium tidal form so that the nature of the (equilibrium) tidal variation with time at a particular point on the earth's surface is not immediately obvious. To analyze the character of this variation, it is convenient to express the zenith angle of the moon in terms of the geographical coordinates θ, ϕ of a point on the earth's surface (θ is colatitude, ϕ is east longitude) and the declination D and west hour angle reckoned from Greenwich α of the moon When this is done, the equilibrium tide can be expressed as the sum of three terms:

$$\bar{\zeta} = \frac{3}{4}\frac{\gamma M}{g}\frac{a^2}{c^3}\Big[(3\sin^2 D - 1)(\cos^2\theta - \tfrac{1}{3}) + \sin 2D \sin 2\theta \cos(\alpha + \phi) + \cos^2 D \sin^2\theta \cos 2(\alpha + \phi)\Big] \qquad (3)$$

where a is the earth's radius

The first term represents a partial tide which is symmetrical about the earth's axis, as it is independent of longitude The only time variation results from the slowly varying lunar declination and distance from earth This tide is called the long-period tide Its actual geographical shape is that of

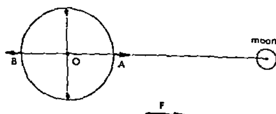


Fig 2. Schematic diagram of the tide-generating force on different points in the earth. The vector sum of this tide-generating force and the constant force F (which does not vary from point to point) gives the force field indicated in Fig. 1 The force F is compensated by the centrifugal force of the earth in its orbital motion.

over into itching Tickle and itch seem also to disappear together in the zone of abnormal sensitivity (secondary hyperalgesia) created as an aftermath of rubbing, pinching, or otherwise overstimulating the cutaneous nerve endings. [F.A.C.]

Tidal bore

A part of a tidal rise in a river which is so rapid that water advances as a wall often several feet high. The phenomenon is favored by a substantial tidal range and a channel which shoals and narrows rapidly upstream, but the conditions are so critical that it is not common. A shoaling channel steepens the tidal curve. If the curve becomes vertical or nearly so, a bore results (Fig. 1). A narrowing channel increases the tidal range. *See* RIVER TIDES. Since the tidal range is greatest at springs, some rivers exhibit bores only at springs (*see* TIDE). While the bore is a very striking feature, Fig. 1 shows that the tide continues to rise after the passage of the bore and that this subsequent rise may be the greater. Bores may be eliminated by changing channel depth or shape.

In North America three bores have been observed: at the head of the Bay of Fundy (Fig. 2), at the head of the Gulf of California, and at the head of Cook Inlet, Alaska. The largest known bore occurs in the Tsientang Kiang, China. At springs it is a wall of water 15 ft high moving upstream a 25 ft/sec. *See* CHANNEL, OPEN. [B.KI.

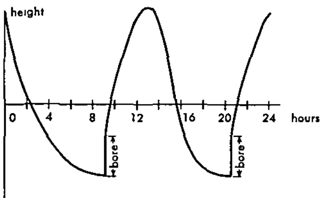


Fig. 1. Tide curve (schematic) of a river with a tidal bore.

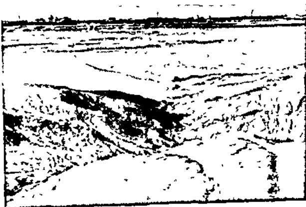


Fig. 2. Tidal bore of the Petitcodiac River, Bay of Fundy, New Brunswick, Canada. Rise of water is about 4 ft. (New Brunswick Travel Bureau)

Tide

The term tide refers to stresses exerted in a bod by the gravitational action of another, and to re lated phenomena resulting from these stresse Every body in the universe raises tides, to some ex tent, on every other. This article deals only wit tides on the earth, since these are fundamentall the same as tides on all bodies. Sometimes varia tions of sea level, whatever their origin, are re ferred to as tides. *See* SEA LEVEL FLUCTUATION

Introduction. The tide-generating forces aris from the gravitational action of sun and moon, th effect of the moon being about twice as effectiv as that of the sun in producing tides. The tidal ef fects of all other bodies on the earth are negligible The tidal forces act to generate stresses in all part of the earth and give rise to relative movements o the matter of the solid earth, ocean, and atmos phere. The earth's rotation gives these movement an alternating character having principal periodic ities of 12.42 and 12.00 hours, corresponding to hal the mean lunar and solar day, respectively.

In the ocean the tidal forces act to generate alternating tidal currents and displacements of the sea surface. These phenomena are important to shipping and have been studied extensively. The main object of tidal studies has been to predict the tidal elevation or current at a given seaport or other place in the ocean at any given time

The prediction problem may be attacked in two ways. Since the relative motions of earth, moon, and sun are known precisely, it is possible to specify the tidal forces over the earth at any past or future time with great precision. It should be possible to relate tidal elevations and currents at any point in the oceans to these forces, making use of classical mechanics and hydrodynamics Such a theoretical approach to tidal prediction has not yet yielded any great success, owing in great part to the complicated shape of the ocean basins.

The other approach, which consists of making use of past observations of the tide at a certain place to predict the tide for the same place, has yielded practical results The method cannot be used for a location where there have been no previous observations Only the frequencies of the many tidal harmonic constituents are derived from knowledge of the movements of earth, moon, and sun. The amplitude and epoch of each constituent are determined from the tidal observations. The actual tide can then be synthesized by summing up an adequate number of harmonic constituents. The method might loosely be thought of as extrapolation.

In the following discussion only the lunar effect is considered, and it is understood that analogous statements apply to the solar effect.

The tide-generating force. If the moon attracted every point within the earth with equal force, there would be no tide. It is the small difference in di-

Table of Darwin's constituents

Constituent	Speed, deg/hour	Coefficient
Long period		
Mf, lunar fortnightly	1 098	157
Ssa, solar semiannual	0 082	.073
Diurnal		
K1, lunisolar	15 041	530
O1, larger lunar	13 943	377
P1, larger solar	14 959	176
Semidiurnal		
M2, principal lunar	28 984	908
S2, principal solar	30 000	423
N2, larger lunar elliptic	28 440	.176
K2, lunisolar	30 082	115

is based on neglect of the inertial forces. These forces are appreciable unless the periods of all free oscillations in the ocean are small compared to those of the tidal forces Actually, there are free oscillations in the ocean (ordinary gravity seiches) having periods of the order of a large fraction of a day, and there may be others (planetary modes) having periods of the order of several days. For

the presence of relatively large meteorological ...ts.

At most places in the ocean and along the coasts, ... level rises and falls in a regular manner The ...est level usually occurs twice in any lunar day,

... range of the tide, is generally greatest near the ...e of full or new moon, and smallest near the ...es of quadrature. This results from the spring-...ap variation in the equilibrium tide The range ... the tide usually exhibits a secondary variation, ...ng greater near the time of perigee (when the ...on is closest to the earth) and smaller at apogee ...hen the moon is farthest away).

The above situation is observed at places where ...e tide is predominantly semidiurnal. At many ...her places, it is observed that one of the two ...ima in any lunar day is higher than the other. ...is effect is known as the diurnal inequality and ...presents the presence of an appreciable diurnal ...tuation. At these places, the tide is said to be of ...e "mixed" type. At a few places, the diurnal tide ...ually predominates, there generally being only ...e high and low tide during the lunar day

Both observation and theory indicate that the ...ean tide can generally be considered linear As a result of this fact, the effect in the ocean of each constituent of series (4) can be considered by itself. Each equilibrium constituent causes a reaction in the ocean. The tide in the ocean is the sum total of all the reactions of the individual constituents. Furthermore, each constituent of the ocean tide is harmonic (sinusoidal) in time. If the amplitude of an equilibrium constituent varies with the nodal cycle of the moon, the amplitude of the oceanic constituent varies proportionately.

As a consequence of the above, the tidal elevation in the ocean can be expressed as

$$\zeta = \Sigma f_i h_i \cos (A_i - G_i) \qquad (5)$$

where $h_i(\theta,\phi)$ is called the amplitude and $G_i(\theta,\phi)$ the Greenwich epoch of each constituent. The summation in (5) extends over all constituents of all species The fs and the As have the same meaning as in expression (4) for the equilibrium tide and are determined from astronomic data.

To specify completely the tidal elevation over the entire surface of the ocean for all time, one would need ocean-wide charts of $h(\theta,\phi)$, called corange charts, and of $G(\theta,\phi)$, called cotidal charts, for each important constituent. Construction of these charts would solve the ultimate problem in tidal prediction. Many attempts have been made to construct cotidal charts, the most notable those of W Whewell, 1833, R. A. Harris, 1904; R Sterneck, 1920, and G. Dietrich, 1944 These attempts have been based on a little theory and far too few observations

Figures 3 and 4 show Dietrich's cotidal chart for M_2. Each curve passes through points having high water at the same time, time being indicated as phase of the M_2 equilibrium argument. A characteristic feature of cotidal charts is the occurrence of points through which all cotidal curves pass. These are called "amphidromic points" Here the amplitude of the constituent under consideration must be zero The existence of such amphidromic points has been borne out by theoretical studies of tides in ocean basins of simple geometric shape. The mechanism which gives rise to amphidromic points is intimately related to the rotation of the earth and the Coriolis force.

The amplitude of a constituent, $h(\theta,\phi)$, is generally high in some large regions of the oceans and low in others, but in addition there are small-scale erratic variations, at least along the coastline. Perhaps this is partly an illusion caused by the placement of some tide gages near the open coast and the placement of others up rivers and estuaries. It is well known that the phase and amplitude of the tide change rapidly as the tidal wave progresses up a river See RIVER TIDES

The range of the ocean tide varies between wide limits The highest range is encountered in the Bay of Fundy, where values exceeding 50 ft have been observed In some places in the Mediterranean, South Pacific, and Arctic, the tidal range never exceeds 2 ft

a spheroid whose axis coincides with the earth's axis and whose oblateness slowly but continuously varies.

The second term of (3) represents a partial tide having, at any instant, maximum elevations at 45°N and 45°S on opposite sides of the earth, and two minimum elevations lying at similar, alternate positions on the same great circle passing through the poles. Because of the factor $\cos(\alpha + \phi)$ the tide rotates in a westerly direction relative to the earth, and any geographical position experiences a complete oscillation in a lunar day, the time taken for α to increase by the amount 2π. Consequently, this partial tide is called the diurnal tide. Because of the factor $\sin 2D$, the diurnal equilibrium tide is zero at the instant the moon crosses the Equator; because of the factor $\sin 2\theta$, there is no diurnal equilibrium tidal fluctuation at the equator nor at the poles.

The third term of (3) is a partial tide having, at any instant, two maximum elevations on the Equator at opposite ends of the earth, separated alternately by two minima also on the Equator. This whole form also rotates westward relatively to the earth, making a complete revolution in a lunar day. But any geographic position on the earth will experience two cycles during this time because of the factor $\cos 2(\alpha + \phi)$. Consequently, this tide is called the semidiurnal tide. Because of the factor $\sin^2 \theta$, there is no semidiurnal equilibrium tidal fluctuation at the poles, while the fluctuation is strongest at the Equator.

It has been found very convenient to consider the equilibrium tide as the sum of a number of terms, called constituents, which have a simple geographical shape and vary harmonically in time. This is the basis of the harmonic development of the tide A great number of tidal phenomena can be adequately described by a linear law, that is, the effect of each harmonic constituent can be superimposed on the effects of the others. Herein is the great advantage of the harmonic method in dealing with tidal problems. The three terms of (3) do not vary with time in a purely harmonic manner. The parameters c and D themselves vary, and the rapidly increasing α does not do so at a constant rate owing to ellipticity and other irregularities of the moon's orbit. Actually, each of the three partial tides can be separated into an entire species of harmonic constituents. The constituents of any one of the three species have the same geographical shape, but different periods, amplitudes, and epochs.

The solar tide is developed in the same way As before, the three species of constituents arise: long-period, diurnal and semidiurnal. The equilibrium tide at any place is the sum of both the lunar and solar tides. When the sun and moon are nearly in the same apparent position in the sky (new moon) or are nearly at opposite positions (full moon), the lunar and solar effects reinforce each other. This condition is called the spring tide. During the spring tide the principal lunar and solar constituents are in phase. At quadrature the solar effect cancels to some extent the lunar effect, the principal lunar and solar constituents being out of phase. This condition is known as the neap tide.

The entire equilibrium tide can now be expressed in the following way:

$$\bar{\zeta} = H[\tfrac{1}{2}(1 - 3\cos^2\theta) \sum_L f_i C_i \cos A_i + \sin 2\theta \sum_D f_i C_i \cos(A_i + \phi) + \sin^2\theta \sum_S f_i C_i \cos(A_i + 2\phi)] \quad (4)$$

where $H = 3\gamma M a^2/g\bar{c}^3 = 54$ cm, and $1/\bar{c}$ represents the mean (in time) value of $1/c$. Each term in the above series represents a constituent. Terms of higher powers of the moon's parallax (a/c) are not included in (4) because of their different latitude dependence, but they are of relatively small importance. The subscripts L, D and S indicate summation over the long-period, diurnal, and semidiurnal constituents, respectively. The Cs are the constituent coefficients and are constant for each constituent They account for the relative strength of all lunar and solar constituents. In a purely harmonic development, such as carried out by A. T. Doodson in 1921, the A parts of the arguments increase linearly with time, and the node factors f are all unity. In George Darwin's "almost harmonic" development of 1882, the constituents undergo a slow change in amplitude and epoch with the 19-year nodal cycle of the moon. The node factors f take this slow variation into account. The As increase almost linearly with time. Tables in U.S. Coast and Geodetic Survey Spec. Publ. 98 enable one to compute the phase of the argument of any of Darwin's constituents at any time, and values of the node factors for each year are given.

In spite of the many advantages of the purely harmonic development, Darwin's method is still used by most agencies engaged in tidal work. In Darwin's classification, each constituent is represented by a symbol with a numerical subscript, 0, 1, or 2, which designates whether the constituent is long-period, diurnal, or semidiurnal. Some of the most important of Darwin's constituents are listed in the table.

The periods of all the semidiurnal constituents are grouped about 12 hours, and the diurnal periods about 24 hours. This results from the fact that the earth rotates much faster than the revolution of the moon about the earth or of the earth about the sun. The principal lunar semidiurnal constituent M_2 beats against the others giving rise to a modulated semidiurnal wave form whose amplitude varies with the moon's phase (the spring-neap effect), distance, and so on. Similarly, the amplitude of the modulated diurnal wave varies with the varying lunar declination, solar declination, and lunar phase. For example, the spring tide at full moon or new moon is manifested by constituents M_2 and S_2 being in phase, thus reinforcing each other. During the neap tide when the moon is at quadrature, the constituents M_2 and S_2 are out of phase, and

Table of Darwin's constituents

Constituent	Speed, deg/hour	Coefficient
Long-period		
Mf, lunar fortnightly	1 098	157
Ssa, solar semiannual	0 082	073
Diurnal		
K_1, lunisolar	15 041	530
O_1, larger lunar	13 943	377
P_1, larger solar	14 959	176
Semidiurnal		
M_2, principal lunar	28 984	908
S_2, principal solar	30 000	.423
N_2, larger lunar elliptic	28 440	176
K_2, lunisolar	30 082	.115

[illegible]

Tides in the ocean. The tide in the ocean deviates markedly from the equilibrium tide, which is not surprising if one recalls that the equilibrium tide is based on neglect of the inertial forces. These forces are appreciable unless the periods of all free oscillations in the ocean are small compared to those of the tidal forces. Actually, there are free oscillations in the ocean (ordinary gravity seiches) having periods of the order of a large fraction of a day and there may be others (planetary modes) having periods of the order of several days. For [illegible]

[illegible] in the presence of relatively large meteorological effects.

At most places in the ocean and along the coasts, sea level rises and falls in a regular manner. The highest level usually occurs twice in any lunar day, the times bearing a constant relationship with the moon's meridional passage. The time between the moon's meridional passage and the next high tide is called the lunitidal interval The difference in level between successive high and low tides, called the range of the tide, is generally greatest near the [illegible]

[illegible] of the tide usually exhibits a secondary variation, being greater near the time of perigee (when the moon is closest to the earth) and smaller at apogee (when the moon is farthest away).

The above situation is observed at places where the tide is predominantly semidiurnal. At many other pla[illegible]

[illegible] presents the presence of an appreciable diurnal variation. At these places, the tide is said to be of the "mixed" type. At a few places, the diurnal tide actually predominates, there generally being only one high and low tide during the lunar day.

Both observation and theory indicate that the ocean tide can generally be considered linear. As a result of this fact, the effect in the ocean of each constituent of series (4) can be considered by itself. Each equilibrium constituent causes a reaction in the ocean. The tide in the ocean is the sum total of all the reactions of the individual constituents. Furthermore, each constituent of the ocean tide is harmonic (sinusoidal) in time. If the amplitude of an equilibrium constituent varies with the nodal cycle of the moon, the amplitude of the oceanic constituent varies proportionately.

As a consequence of the above, the tidal elevation in the ocean can be expressed as

$$\zeta = \Sigma f_i h_i \cos (A_i - G_i) \qquad (5)$$

where $h_i(\theta,\phi)$ is called the amplitude and $G_i(\theta,\phi)$ the Greenwich epoch of each constituent The summation in (5) extends over all constituents of all species. The fs and the As have the same meaning as in expression (4) for the equilibrium tide and are determined from astronomic data.

To specify completely the tidal elevation over the entire surface of the ocean for all time, one would need ocean-wide charts of $h(\theta,\phi)$, called corange charts, and of $G(\theta,\phi)$, called cotidal charts, for each important constituent. Construction of these charts would solve the ultimate problem in tidal prediction Many attempts have been made to construct cotidal charts, the most notable those of W Whewell, 1833; R. A. Harris, 1904; R Sterneck, 1920; and G Dietrich, 1944 These attempts have been based on a little theory and far too few observations

Figures 3 and 4 show Dietrich's cotidal chart for M_2 Each curve passes through points having high water at the same time, time being indicated as phase of the M_2 equilibrium argument. A characteristic feature of cotidal charts is the occurrence of points through which all cotidal curves pass. These are called "amphidromic points." Here the amplitude of the constituent under consideration must be zero The existence of such amphidromic points has been borne out by theoretical studies of tides in ocean basins of simple geometric shape. The mechanism which gives rise to amphidromic points is intimately related to the rotation of the earth and the Coriolis force.

The amplitude of a constituent, $h(\theta,\phi)$, is generally high in some large regions of the oceans and low in others, but in addition there are small-scale erratic variations, at least along the coastline. Perhaps this is partly an illusion caused by the placement of some tide gages near the open coast and the placement of others up rivers and estuaries. It is well known that the phase and amplitude of the tide change rapidly as the tidal wave progresses up [illegible]

[illegible] of Fundy, where values exceeding 50 ft have been observed In some places in the Mediterranean, South Pacific, and Arctic, the tidal range never exceeds 2 ft.

a spheroid whose axis coincides with the earth's axis and whose oblateness slowly but *continuously* varies.

The second term of (3) represents a partial tide having, at any instant, maximum elevations at 45°N and 45°S on opposite sides of the earth, and two minimum elevations lying at similar, alternate positions on the same great circle passing through the poles. Because of the factor cos $(\alpha + \phi)$ the tide rotates in a westerly direction relative to the earth, and any geographical position experiences a complete oscillation in a lunar day, the time taken for α to increase by the amount 2π. Consequently, this partial tide is called the diurnal tide. Because of the factor sin $2D$, the diurnal equilibrium tide is *zero at the instant* the moon crosses the Equator; *because of the factor sin* 2θ, there is no diurnal equilibrium tidal fluctuation at the equator nor at the poles.

The *third term of* (3) is a *partial tide* having, at any instant, two maximum elevations on the Equator at opposite ends of the earth, separated alternately by two minima also on the Equator. This whole form also rotates westward relatively to the earth, making a complete revolution in a lunar day. But any geographic position on the earth will experience two cycles during this time because of the factor cos $2(\alpha + \phi)$. Consequently, this tide is called the semidiurnal tide. Because of the factor $\sin^2 \theta$, there is no semidiurnal equilibrium tidal fluctuation at the poles, while the fluctuation is strongest at the Equator.

It has been found very convenient to consider the *equilibrium tide* as the sum of a number of terms, called constituents, which have a simple geographical shape and vary harmonically in time. This is the basis of the harmonic development of the tide. A great number of tidal phenomena can be adequately described by a linear law; that is, the effect of each harmonic constituent can be superimposed on the effects of the others. Herein is the great advantage of the harmonic method in dealing with tidal problems. The three terms of (3) do not vary with time in a purely harmonic manner. The parameters c and D themselves vary, and the rapidly increasing α does not do so at a constant rate owing to ellipticity and other irregularities of the moon's orbit. Actually, each of the three partial tides can be separated into an entire species of harmonic constituents. The constituents of any one of the three species have the same geographical shape, but different periods, amplitudes, and epochs.

The solar tide is developed in the same way. As before, the three species of constituents arise: long-period, diurnal and semidiurnal. The equilibrium tide at any place is the sum of both the lunar and solar tides. When the sun and moon are nearly in the same apparent position in the sky (new moon) or are nearly at opposite positions (full moon), the lunar and solar effects reinforce each other. This condition is called the spring tide. During the spring tide the principal lunar and solar constituents are in phase. At quadrature the solar effect cancels to some extent the lunar effect, the principal lunar and solar constituents being out of phase. This condition is known as the neap tide.

The entire equilibrium tide can now be expressed in the following way:

$$\bar{\zeta} = H[\tfrac{1}{2}(1 - 3\cos^2\theta) \sum_L f_i C_i \cos A_i + \sin 2\theta \sum_D f_i C_i \cos (A_i + \phi) + \sin^2\theta \sum_S f_i C_i \cos (A_i + 2\phi)] \quad (4)$$

where $H = 3\gamma M a^2/g\bar{c}^3 = 54$ cm, and $1/\bar{c}$ represents the mean (in time) value of $1/c$. Each term in the above series represents a constituent. Terms of higher powers of the moon's parallax (a/c) are not included in (4) because of their different latitude dependence, but they are of relatively small importance. The subscripts L, D and S indicate summation over the long-period, diurnal, and semidiurnal constituents, respectively. The *Cs are the* constituent coefficients and are constant for each constituent. They account for the relative *strength* of all lunar and solar constituents. In a purely harmonic development, such as carried out by A. T. Doodson in 1921, the A parts of the arguments increase linearly with time, and the node factors f are all unity. In George Darwin's "almost harmonic" development of 1882, the constituents undergo a slow change in amplitude and epoch with the 19-year nodal cycle of the moon. The node factors f take this slow variation into account. The As increase almost linearly with time. Tables in U.S. Coast and Geodetic Survey Spec. Publ. 98 enable one to compute the phase of the argument of any of Darwin's constituents at any time, and values of the node factors for each year are given.

In spite of the many advantages of the purely harmonic development, Darwin's method is still used by most agencies engaged in tidal work. In Darwin's classification, each constituent is represented by a symbol with a numerical subscript, 0, 1, or 2, which designates whether the constituent is long-period, diurnal, or semidiurnal. Some of the most important of Darwin's constituents are listed in the table.

The periods of all the semidiurnal constituents are grouped about 12 hours, and the diurnal periods about 24 hours. This results from the fact that the earth rotates much faster than the revolution of the moon about the earth or of the earth about the sun. The principal lunar semidiurnal constituent M_2 beats against the others giving rise to a modulated semidiurnal wave form whose amplitude varies with the moon's phase (the spring-neap effect), distance, and so on. Similarly, the amplitude of the modulated diurnal wave varies with the varying lunar declination, solar declination, and lunar phase. For example, the spring tide at full moon or new moon is manifested by constituents M_2 and S_2 being in phase, thus reinforcing each other. During the neap tide when the moon is at quadrature, the constituents M_2 and S_2 are out of phase, and

tion. The variation of velocity with depth due to the stratification of the water is associated with internal wave motion. Serial observations made from anchored or drifting ships have disclosed prominent tidal periodicities in the vertical thermal structure of the water. *See* WAVE (INTERNAL).

Dynamics of the ocean tide. The theoretical methods for studying tidal dynamics in the oceans were put forth by Laplace in the eighteenth century. The following assumptions are introduced: (1) the water is homogeneous; (2) vertical displacements and velocities of the water particles are small in comparison to the horizontal displacements and velocities, (3) the water pressure at any point in the water is given adequately by the hydrostatic law; that is, is equal to the head of water above the given point; (4) all dissipative forces are neglected; (5) the ocean basins are assumed rigid (as if there were no bodily tide), and the gravitational potential of the tidally displaced masses is neglected; and (6) the tidal elevation is small compared to the water depth.

If assumptions (1) and (3) are valid, it can readily be shown that the tidal currents are uniform with depth. This is a conclusion which is not in complete harmony with observations, and there

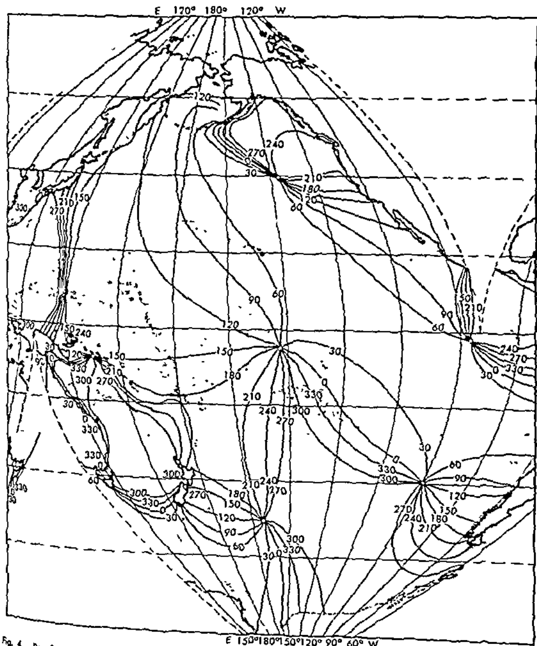
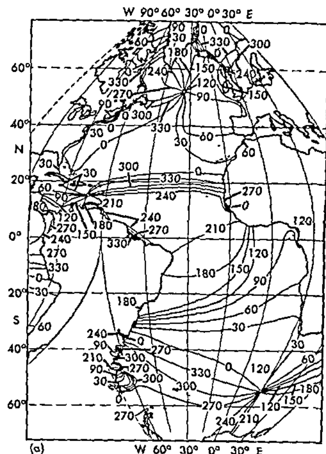


Fig. 4 Pacific Ocean cotidal chart for M_2. (According to G Dietrich, from Veroffentl Inst Meeresk, n.s A, Geograph-naturw Reihe, no. 41, 1944)



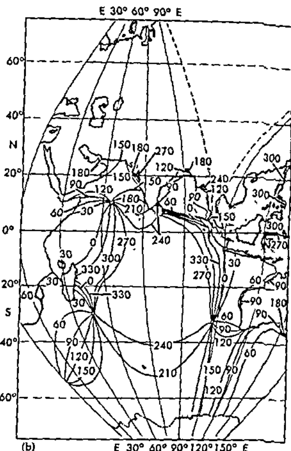


Fig. 3. *Cotidal chart for M_2. (a) Atlantic Ocean (b) Indian Ocean. (According to G Dietrich, from Veroffentl. Inst Meeresk., n s. A, Geograph-naturw Reihe, no. 41, 1944)*

The tide may be considerably different in small adjacent seas than in the nearby ocean, and here resonance phenomena frequently occur The periods of free oscillation of a body of water are determined by their boundary and depth configuration (*see* SEICHE). If one of these free periods is near that of a large tidal constituent, the latter may be amplified considerably in the small sea. The large tidal range in the Bay of Fundy is an example of this effect. Here the resonance period is nearly 12 hours, and it is the semidiurnal constituents that are large. The diurnal constituents are not extremely greater in the Bay of Fundy than in the nearby ocean.

In lakes and other completely enclosed bodies of water the periods of free oscillation are usually much smaller than those of the tidal constituents Therefore the tide in these places obeys the principles of statics. Since there is no tidal variation in the total volume of water in lakes the mean surface elevation does not change with the tide The surface slope is determined by the slope of the equilibrium tide, and the related changes in elevation are usually very small, of the order of a fraction of a millimeter for small lakes.

Tidal currents. The south and east components of the tidal current can be developed in the same way as the tidal elevation since they also depend linearly on the tidal forces. Consequently, the same analysis and prediction methods can be used. Expressions similar to (5) represent the current components, each constituent having its own amplitude and phase at each geographic point. It should be emphasized that the current speed or direction cannot be developed in this way since these are not linearly related to the tidal forces.

Only in special cases are the two tidal current components exactly in or out of phase, and so the tidal current in the ocean is generally rotatory. A drogue or other floating object describes a trajectory similar in form to a Lissajous figure. In a narrow channel only the component along its axis is of interest. Where shipping is important through such a channel or port entrance, current predictions, as well as tidal height predictions, are sometimes prepared.

Owing to the rotation of the earth, there is a gyroscopic, or Coriolis, force acting perpendicularly to the motion of any water particle in motion. In the Northern Hemisphere this force is to the right of the current vector. The horizontal, or tractive, component of the tidal force generally rotates in the clockwise sense in the Northern Hemisphere As a result of both these influences the tidal currents in the open ocean generally rotate in the clockwise sense in the Northern Hemisphere, and in the counterclockwise sense in the Southern Hemisphere There are exceptions, however, and the complete dynamics should be taken into account. *See* CORIOLIS ACCELERATION AND FORCE.

The variation of the tidal current with depth is not well known. It is generally agreed that the current would be constant from top to bottom were it not for stratification of the water and bottom fric-

factor is of the order of several tens of microgals.

Atmospheric tides. Since air, as other matter, is subject to gravitational influence, there are tides in the atmosphere possessing many features of similarity with those in the ocean. One of the characteristics of these tides is a small oscillatory variation in the atmospheric pressure at any place. This fluctuation of pressure, as in the case of the ocean tide, may be considered as the sum of the usual tidal constituents, and standard tidal analysis and prediction methods may be used. The principal lunar semidiurnal constituent M_2 of the pressure variation has been determined for a number of places, and found to have an amplitude of the order of 0.03 millibars. The dynamical theory of these tides has been the subject of considerable study. The equations which have been considered have the same general form as those for ocean tides. The S_2 constituent shows a much larger oscillation with an amplitude of the order of 1 millibar, but here diurnal heating dominates the gravitational effects. If diurnal heating were the whole story one would expect an even larger S_1 effect, and the fact that S_2 is larger is attributed to an atmospheric resonance near 12 hours.

Tidal analysis and prediction. The distribution in space and time of the tidal forces within the earth is precisely known from astronomic data. The effects of these forces on the oceans cannot, by present methods, be described in detail on a worldwide basis because of the difficult nature of the dynamical relationships and the complicated shape of the ocean basins. Practical prediction methods make use of past observations at the place under consideration [illegible]

[illegible] ponents of the tidal current, and so on—which depends linearly on the tidal forces. The frequencies, or periods, of the tidal constituents are determined by the astronomic data, and the harmonic [illegible] btained [illegible] consideration, where the values of h are the amplitudes of whatever tidal variable is being predicted. In this discussion the sea-level elevation will be used as an example, since it is the variable for which predictions are most commonly made.

Tidal analysis consists of determining the harmonic constants from a record of sea level at a given place. The procedure is basically the same for each constituent, but is most easily described for the series of constituents, S_1, S_2, S_3, . . . , whose periods are submultiples of 24 hours. Suppose that the tidal elevation at 1:00 A.M. is averaged for all the days of the tide record, and similarly for 2:00 A.M., 3:00 A.M., and for each hour of the day. The 24 values thus obtained represent the average diurnal variation during the entire record. Any constituent whose period is not a submultiple of 24 hours will contribute very little to the average of all the 1:00 A.M. values since its phase will be different from one day to the next, and its average value at 1:00 A.M. will be very close to zero for a long record. The same is true for each hour of the day, and so its average diurnal variation is small. The longer the record the freer will be the average diurnal oscillation from the effects of the other constituents. The diurnal oscillation is then analyzed by the well known methods of harmonic analysis to determine the amplitudes and phases of all the harmonics of the 24-hour oscillation. See FOURIER SERIES AND INTEGRALS.

The same procedure is used for each other constituent; that is, the tide record is divided into consecutive constituent days, each equal to the period (or double the period in the case of the semidiurnal constituents) of the constituent. If the tide record is [illegible] solar hour, there is a slight com- [illegible] is overcome by substituting the [illegible] nearest the required time and later compensating the consistent error introduced by an augmenting factor.

Since the record length is always finite, the harmonic constants of a constituent determined by this method are somewhat contaminated by the effects of other constituents. A first-order correction of these effects can be made by an elimination procedure. In general it is more efficient to take the record length equal to the synodic (beat) period of two or more of the principal constituents. Of course, the longer the record the better. Standard analyses consist of 29 days, 58 days, 369 days, and so on.

It is not practical to determine the harmonic constants of the lesser constituents in this way if errors or uncertainties of the data are of the same order of magnitude as their amplitudes. If tidal oscillations in the oceans were far from resonance then we should expect the amplitude of each constituent to be approximately proportional to its theoretical coefficient, and the local epochs all to be near the same value. In other words, for the semidiurnal constituent X, we should expect that

$$\frac{H(X)}{C(X)} = \frac{H(M_2)}{C(M_2)} \qquad G(X) = G(M_2) \qquad (8)$$

Here X is referred to M_2 for the reason that the latter is one of the principal constituents, whose harmonic constants can be determined with best accuracy. Any other important constituent could be used. Inferring the harmonic constants of the lesser constituents by means of Eq. (8) is sometimes preferable to direct means. It should be borne in mind that a constituent of one species cannot be inferred from one of another species because their equilibrium counterparts have different geographic shapes and no general relationship such as (8) exists.

Once the harmonic constants are determined the tide is synthesized according to Eq. (5) usually with the help of a special "tide-predicting ma-

are internal wave modes thus left out of Laplace's theory. Nevertheless the main features of the tide are probably contained in the equations.

The water motion in the oceans is, in theory, determined by knowledge of the shape of the ocean basins and the tide-generating force (or equilibrium tide) at every point in the oceans for all time. The theory makes use of two relations: (1) the equation of continuity, which states that the rate of change of water mass in any vertical column in the ocean is equal to the rate at which water is flowing into the column; and (2) the equations of motion, which state that the total acceleration of a water "particle" (relative to an inertial system, thus taking into account the rotation of the earth) is equal to the total force per unit mass acting on that particle. Under the above assumptions, the equation of continuity takes the form

$$\frac{\partial \zeta}{\partial t} = -\frac{1}{a \sin \theta}\left[\frac{\partial}{\partial \theta}(ud \sin \theta) + \frac{\partial}{\partial \phi}(vd)\right] \quad (6)$$

where $d(\theta,\phi)$ is the water depth. The equations of motion in the southward and eastward directions, respectively, are given by

$$\frac{\partial u}{\partial t} - 2\omega v \cos \theta = -\frac{g}{a}\frac{\partial}{\partial \theta}(\zeta - \bar{\zeta})$$

$$\frac{\partial v}{\partial t} + 2\omega u \cos \theta = -\frac{g}{a}\csc \theta \frac{\partial}{\partial \phi}(\zeta - \bar{\zeta}) \quad (7)$$

where ω designates the angular rate of rotation of the earth, u and v the south and east components on the tidal current. All other quantities are as previously defined.

It is probable that exact mathematical solutions to the above equations, taking even approximately into account the complicated shape of the ocean basins, will never be obtained. However, the equations have certain features which serve to give us some insight into the nature of ocean tides For instance it is evident that if many equilibrium tides are acting simultaneously on the ocean, then the ocean tide will be the sum of the individual reactions. This linearity results directly from above assumption (6). In certain shallow regions of the ocean the tides are noticeably distorted, as would be expected if assumption (6) were violated This distortion is usually considered as resulting from the presence of so-called shallow-water constituents having frequencies equal to harmonics and to beat frequencies of the equilibrium constituents. These must be considered, at some places, or there will be large discrepancies between prediction and observation (*see* RIVER TIDES). Certain mathematical solutions to equations (6) and (7) have been obtained for hypothetical ocean basins of simple geometric shape. Laplace solved them for an ocean of constant depth covering the entire earth. Several solutions have been obtained for an ocean of constant depth bounded by two meridians. The result of one of the solutions obtained by J. Proudman and A. Doodson is shown in Fig. 5, which represents a cotidal chart of the K_2 tide in an ocean of

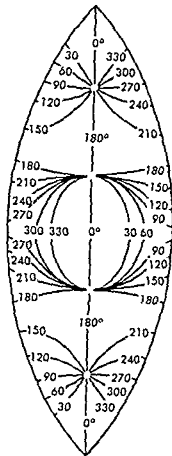


Fig. 5. Cotidal chart for K_2 in a hypothetical ocean of constant depth bounded by meridians 70° apart. (According to J. Proudman and A. T. Doodson, from A. T. Doodson and H. D. Warburg, *Admiralty Manual of Tides*, London, 1941)

depth 14,520 ft bounded by meridians 70° apart. The K_2 tide was calculated because of mathematical simplifications, but the M_2 tide should be quite similar Comparison of Fig. 5 with the Atlantic Ocean in Fig 3 discloses no striking similarities except for the general occurrence of amphidromic systems

The bodily tide. The solid part of the earth suffers periodic deformation resulting from the tide-generating forces just as the oceans do *See* EARTH TIDES

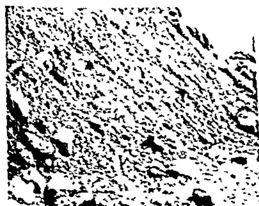
The gravest known modes of free oscillation of the solid earth have periods of the order of an hour, much shorter than those of the principal tidal constituents. Therefore, the principles of statics can be used to describe the bodily tide, in contrast with tides in the oceans and atmosphere, where the inertial effect is important.

Associated with the bodily tide are periodic changes in gravity, manifesting themselves as (1) a variation of the vertical, or plumb line, with respect to any solid structure imbedded in the earth's crust, and (2) a variation in the magnitude of the acceleration of gravity at any point. These effects arise from the gravitational attraction of the tidally displaced matter of the earth (solid, ocean, and atmosphere) as well as directly from the tide-generating forces. The magnitude of the former

Till

The unstratified portion of glacial drift. The unsorted materials of the till are deposited by the advancing ice, or as a result of melting or evaporation of the ice during the waning stage of glaciation. The term boulder clay refers to a common variety of till containing embedded particles ranging in size from fine grains to boulders.

The texture of till is characterized by extreme variation in grain size. The matrix consists of the finer clastic materials, clay, silt, and sand. Randomly embedded in this are larger fragments including boulders of many cubic yards in volume. The coarser fragments, cobbles and boulders, commonly display faceting and striations caused by abrasion during transport by the ice. Careful study may reveal a preferred orientation of the larger fragments. This is usually the only indication of stratification. Lenses of stratified sand, gravel, or silt which occur locally within the till represent the local action of melt water.

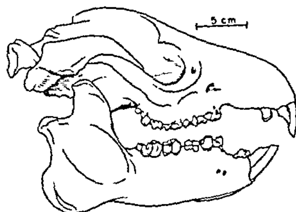


Exposure of glacial till at the Black Rocks near Llandudno, Wales. Heterogeneous debris, ranging in size from large boulders to fine powder, displays no assortment or stratification (Photograph by K F Mather)

Till consists of physically broken and disintegrated, but essentially undecomposed, rock and mineral fragments. Commonly all types of rocks are represented, but igneous and metamorphic ones predominate. These materials, under favorable climatic conditions, are readily converted into excellent soils. *See* GLACIATED TERRANE. [C.J R]

Tillodontia

These extinct quadrupedal land mammals are known from lower Tertiary deposits in the Northern Hemisphere. They progressively developed enlarged rootless, second incisors that are remarkably rodentlike. Their cheek teeth are blunt-cusped with basically three principal cusps in uppers and five in lowers. The principal cusp on the inside of the upper teeth and the cusps on the outside of the lower teeth tend to form as curved columnar structures, giving the unworn tillodont tooth a unique



Side view of the cranium and lower jaw of a middle Eocene tillodont, *Trogosus* (After C. L. Gazin, 1953)

appearance as compared to other mammals. Tillodont feet have five clawed toes.

Esthonyx, a small form with rooted incisors of the late Paleocene and early Eocene of North America and Europe, is apparently ancestral to the large *Trogosus* and *Tillodon* of the middle Eocene of North America. *Adapidium* of the middle or late Eocene of China is of uncertain subordinal affinity.

C. L. Gazin, the principal recent worker on the group, concludes that tillodonts may have had a nearly common origin with the order Pantodonta in a Paleocene arctocyonid creodont stock but does not rule out the possibility that they might have stemmed from the Insectivora. *See* CARNIVORA FOSSILS; INSECTIVORA FOSSILS; PANTODONTA. [D.E.S.]

Time

The dimension of the physical universe which, at a given place, orders the sequence of events, also, a designated instant in this sequence, as the time of day, technically known as an epoch.

Measurement. Time measurement consists in *counting the repetitions* of any recurring phenomenon, and if the interval between successive recurrences is sensible, in subdividing it. The phenomenon most used has been the rotation of Earth, where the counting is by days. Days are measured by observing the meridian passages of stars, and are subdivided with the aid of precision clocks. The day is, however, subject to variations in duration; consequently, when the utmost precision is required, years are measured and subdivided in preference to days (*see* DAY).

A *determination* of time is synonymous with the establishment of an epoch; it consists in ascertaining the clock correction, which is the correction that should be applied to the reading of a clock (positive if the clock is slow) at a specified epoch. A time interval may be measured in two ways: as the duration between two known epochs, or simply by counting from an arbitrary starting point, as is done with a stop watch.

Time units are the intervals between successive *recurrences of phenomena*, as the period of rotation of Earth, also arbitrary multiples and subdivi-

chine," although any means of computation could be used. Usually only the times and heights of high and low water are published in the predictions.

Tidal friction. The dissipation of energy by the tide is important in the study of planet motion because it is a mechanism whereby angular momentum can be transferred from one type of motion to another. An appreciable amount of tidal dissipation takes place in the ocean, and possibly also in the solid earth. In 1952 Sir Harold Jeffreys estimated that about half the tidal energy present in the ocean at any time is dissipated each day. A large part of this dissipation takes place by friction of tidal currents along the bottom of shallow seas and shelves and along the coasts. The rate of dissipation is so large that there should be a noticeable effect on the tide in the oceans.

If the planet's speed of rotation is greater than its satellite's speed of revolution about it, as is the case in the earth-moon system, then tidal dissipation always tends to decelerate the planet's rotation, with the satellite's speed of revolution changing to conserve angular momentum of the entire system. The moon's attraction on the irregularly shaped tidal bulge on the earth exerts on it a decelerating torque. Thus tidal friction tends to increase the length of day, to increase the distance between earth and moon, and to increase the lunar month, but these increases are infinitesimal. The day may have lengthened by 1 second during the last 120,000 years because of tidal friction and other factors. [G.W.G.]

Bibliography: A. Defant, *Ebb and Flow. The Tides of Earth, Air, and Water*, 1958; H. Jeffreys, *The Earth*, 1952, H. Lamb, *Hydrodynamics*, 6th ed., 1945; H. A. Marmer, *The Tide*, 1926; P. Schureman, *A Manual of Harmonic Analysis and Prediction of Tides*, U.S. Coast and Geodetic Survey, Spec. Publ. 98, 1941.

Tie rod

A tie rod or tie bar, usually circular in cross section, is used in structural parts of machines to tie together or brace connected members, or in moving parts of machines or mechanisms it may connect arms or parts to transmit motion. In the first use the rod ends are usually a threaded fastening, while in the latter they are usually forged into an eye for a pin connection.

In steering systems of automotive vehicles, the rod connects the arms of steering knuckles of each wheel. The connection between the rod and arms is a ball and socket joint.

In pressure piping, large forces are produced between connected parts. The pipes or parts are constrained by tie rods that may be rectangular in cross section, with pinned ends. [P.H.B.]

Tiger

A large carnivore, *Felis tigris*, of the family Felidae, occurring from Siberia southward through India, the Malay Peninsula, Java, and Sumatra. The Siberian tiger is the largest, attaining a length of 13 ft and weighing as much as 650 lb. The *Bengal tiger* is smaller and more brilliantly colored.



Bengal tiger, *Felis tigris*; length 6½ ft. *(From P. Martin Duncan, ed., Cassell's Natural History, Cassell)*

In parts of India, tigers take many human lives each year and destroy thousands of domestic animals. Not all tigers kill humans. Some individuals, notably old females, may turn killer as a result of hunger when they are too old to catch more elusive animals. The brilliant striping of the tiger is effective concealment in the animal's home environment. See CARNIVORA. [J.D.B.]

Tile

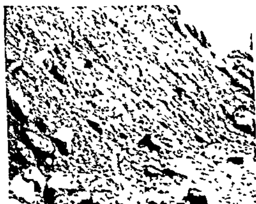
A glazed or unglazed ceramic building unit of thin cross section, used for surface treatment of roofs, walls, and floors. By extension, the term is now applied to other units of similar shape and use but of different materials, such as asphalt, cork, linoleum, vinyl and vinyl asbestos, and porcelain enamel. Clay cast in the form of hollow blocks and either unglazed or glazed is called structural clay tile. Made into pipe, unglazed clay tile is called drain tile. See CLAY PRODUCTS, ARCHITECTURAL.

Glazed roofing tiles of varying colors and shapes are now limited in use in the Occident mainly to churches, other public buildings, and the homes of the wealthy. Wall tiles (fired with a vitreous glaze), used since early Egyptian days as an ornamental and durable wall surfacing, are now most commonly used in bathrooms and kitchens, and also, mainly in nonresidential buildings, for fireproof and easily cleaned corridor walls and building exteriors. Modern floor tiles are of two types, (1) ceramic mosaic, quarry, paver, or slab (integrally colored and glazed or unglazed), and (2) crystalline (colored on the surface only). They are used for bathroom floors and swimming pools, in the subtropics as a substitute for wood floors, and very recently as art mosaic for richly designed walls, indoors or out. Drain tiles today are made of concrete as well as of clay. [G.C.O.]

Till

The unstratified portion of glacial drift. The unsorted materials of the till are deposited by the advancing ice, or as a result of melting or evaporation of the ice during the waning stage of glaciation. The term boulder clay refers to a common variety of till containing embedded particles ranging in size from fine grains to boulders.

The texture of till is characterized by extreme variation in grain size. The matrix consists of the finer clastic materials, clay, silt, and sand. Randomly embedded in this are larger fragments including boulders of many cubic yards in volume. The coarser fragments, cobbles and boulders, commonly display faceting and striations caused by abrasion during transport by the ice. Careful study may reveal a preferred orientation of the larger fragments. This is usually the only indication of stratification. Lenses of stratified sand, gravel, or silt which occur locally within the till represent the local action of melt water.

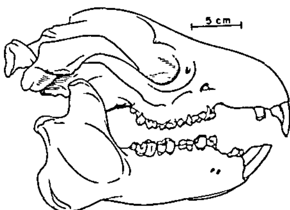


Exposure of glacial till at the Black Rocks near Llandudno, Wales. Heterogeneous debris, ranging in size from large boulders to fine powder, displays no assortment or stratification. (*Photograph by K. F. Mather*)

Till consists of physically broken and disintegrated, but essentially undecomposed, rock and mineral fragments. Commonly all types of rocks are represented, but igneous and metamorphic ones predominate. These materials, under favorable climatic conditions, are readily converted into excellent soils. See GLACIATED TERRANE. [C.J.R.]

Tillodontia

These extinct quadrupedal land mammals are known from lower Tertiary deposits in the Northern Hemisphere. They progressively developed enlarged, rootless, second incisors that are remarkably rodentlike. Their cheek teeth are blunt-cusped with basically three principal cusps in uppers and five in lowers. The principal cusp on the inside of the upper teeth and the cusps on the outside of the lower teeth tend to form as curved columnar structures, giving the unworn tillodont tooth a unique



Side view of the cranium and lower jaw of a middle Eocene tillodont, *Trogosus*. (*After C. L. Gazin, 1953*)

appearance as compared to other mammals. Tillodont feet have five clawed toes.

Esthonyx, a small form with rooted incisors of the late Paleocene and early Eocene of North America and Europe, is apparently ancestral to the large *Trogosus* and *Tillodon* of the middle Eocene of North America. *Adapidium* of the middle or late Eocene of China is of uncertain subordinal affinity.

C. L. Gazin, the principal recent worker on the group, concludes that tillodonts may have had a nearly common origin with the order Pantodonta in a Paleocene arctocyonid creodont stock but does not rule out the possibility that they might have stemmed from the Insectivora. See CARNIVORA FOSSILS; INSECTIVORA FOSSILS; PANTODONTA. [D.E.S.]

Time

The dimension of the physical universe which, at a given place, orders the sequence of events; also, a designated instant in this sequence, as the time of day, technically known as an epoch.

Measurement. Time measurement consists in counting the repetitions of any recurring phenomenon, and if the interval between successive recurrences is sensible, in subdividing it. The phenomenon most used has been the rotation of Earth, where the counting is by days. Days are measured by observing the meridian passages of stars, and are subdivided with the aid of precision clocks. The day is, however, subject to variations in duration; consequently, when the utmost precision is required, years are measured and subdivided in preference to days (see DAY).

A determination of time is synonymous with the establishment of an epoch; it consists in ascertaining the clock correction, which is the correction that should be applied to the reading of a clock (positive if the clock is slow) at a specified epoch. A time interval may be measured in two ways: as the duration between two known epochs, or simply by counting from an arbitrary starting point, as is done with a stop watch.

Time units are the intervals between successive recurrences of phenomena, as the period of rotation of Earth, also arbitrary multiples and subdivi-

sions of these intervals, such as the hour, being one twenty-fourth of a day, and the minute, being one-sixtieth of an hour. *See* MONTH; SECOND (TIME UNIT); YEAR.

Time bases. Several phenomena are used as the time base to be divided into hours. For astronomical purposes, sidereal time is used; for terrestrial purposes, solar time is used.

Sidereal time. The hour angle of the vernal equinox is the measure of sidereal time. It is reckoned from 0 to 24 hr, which are subdivided into 60 sidereal min and the minutes into 60 sidereal sec.

Sidereal clocks are used for convenience in most astronomical observatories, because a star, or other object outside the solar system, comes to the same place in the sky each night at virtually the same sidereal time.

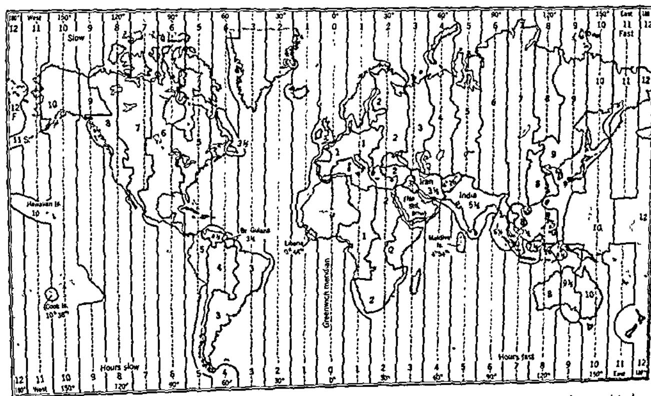
Solar time The hour angle of the Sun is the apparent solar time. The only true indicator of apparent solar time is a sundial.

Mean solar time has been devised to eliminate the irregularities in apparent solar time that arise from the obliquity of the ecliptic and the varying speed of Earth in its orbit around the Sun. It is the hour angle of a fictitious mean Sun, an imagined point moving uniformly along the celestial equator at the same rate as the average rate of the actual Sun along the ecliptic. In practice, it is intervals of sidereal time that are directly observed, and afterward converted into intervals of mean solar time by division by 1 00273790926

Because sidereal and solar time are both defined as hour angles, at any instant, they vary from place to place on Earth. When the mean Sun is on the meridian of Greenwich, the mean solar time is 12 noon at Greenwich. At that instant of absolute time, the mean solar time for all places west of Greenwich is earlier than noon and for all places east of Greenwich later than noon, the progression being at the rate of 1 hr for each 15° of longitude, or 12 hr for a semicircumference. Thus, at the same instant, at a short distance east of the 180th meridian, the mean solar time is 12:01 A.M., and at a short distance west of the 180th meridian it is 11:59 P.M. of the same day. Thus, persons going westward around the earth must advance their time one day, and those going eastward must retard their time one day, in order to be in agreement with their neighbors when they return home. The *International Date Line* is the name given to a line following approximately the 180th meridian, but avoiding inhabited islands, where the change of date is made. Mean solar time at Greenwich is called Greenwich mean time.

Zone and standard times. To avoid the inconvenience of the continuous change of mean solar time with longitude, zone time or civil time is the time generally used. Earth is divided into 24 time zones, each approximately 15° wide and centered on standard longitudes 0°, 15°, 30°, and so on as illustrated. Within each of these zones, the time kept is the mean solar time of the standard meridian. Most civilized nations use zone time.

Zone time is reckoned from 0 to 24 hr for most official purposes, the time in hours and minutes being expressed by a four-figure group followed by the zone designation, as 1009 zone plus 5, referring to the zone 75° west of Greenwich The various zones are sometimes designated by letters, especially the Greenwich zone, which is Z, 1509 Z mean-



World is divided into 24 standard time zones, each differing from Greenwich Civil Time an additional hour. Some countries use half-hour intervals or fractional hours, Venezuela being 4½ hours slow and India being 5½ hours fast on Greenwich Civil Time

g 1509 Greenwich mean time. The zone centered on the 180th meridian is divided into two parts, the one east of the date line being designated plus 12 and the other minus 12. The time July 2, 2400 is identical with July 3, 0000.

In civil life, the designations A.M. and P.M. are often used, usually with punctuation between hours and minutes, thus 1009 may be written as 10:09 A.M. and 1509 as 3 09 P.M. In this system, noon is correctly designated as 12:00 M. Sometimes, July 2, 2400 is called July 2, 12:00 P.M. The designation July 3, 12.00 A.M. is not used, although it is logically the same as July 2, 12·00 P.M. The designations for noon and midnight are, however, often confused, and it is better to write 12:00 noon and July 2–3, 12.00 midnight in order to avoid ambiguity. In some occupations where time is of special importance, there is a rule against using 12.00 at all, 11:59 or 12.01 being substituted The time 1 min after midnight is 12·01 A.M and 1 min after noon is 12 01 P.M.

The figure shows the designations of the various time zones, the longitudes of the standard meridians, the letter designations, and the times in the various zones when it is noon at Greenwich

In the United States, the boundaries of the time zones are fixed by the Interstate Commerce Commission, and here as elsewhere, the actual boundaries depart considerably from the meridians exactly midway between the standard meridians

Ships at sea and transoceanic planes use Greenwich mean time for navigation and communications, but for regulating daily activities on board, they use any convenient approximation to zone time, avoiding frequent changes during daylight hours. [G M C.]

Daylight saving time Large sections of the United States and many European countries set their time one hour ahead during summer months, into daylight saving time, or summer time Thus 6 A.M. standard time becomes 7 A M daylight sav-

on electric power and the residents can benefit from the daylight hour in the evening. This prac-

equal.

The United States has used daylight saving time, country wide, in both world wars, and in the last war it held through summer and winter from February, 1942, to October, 1945 England moved her clocks 2 hr ahead of Greenwich civil time during the last war. [V H E]

Ephemeris time. The orbital motions of the moon and planets are used for ephemeris time It is free from the irregularities in mean solar time caused by variations in the rate of rotation of the earth, and is determined in practice as a correction to Greenwich mean time, which will bring observations of the right ascension and declination of the moon into agreement with the theoretically calculated values. Clocks are not actually regulated to ephemeris time; it is sufficient to keep a record of the corrections necessary.

Time signals are pips emitted by radio stations in most civilized countries, enabling the listener to ascertain the zone time accurate to a small fraction of a second. In the United States time signals consist of coded seconds pulses emitted at frequent intervals by naval radio stations and by station WWV of the National Bureau of Standards, the precision being controlled by the U.S. Naval Observatory. *See* CALENDAR. [G.M.C.]

Time constant

The time required for a physical quantity to change its initial (zero-time) magnitude by the factor $(1 - 1/\epsilon)$ when the physical quantity is varying as a function of time, $f(t)$, according to the decreasing exponential function (Fig 1)

$$f(t) = \epsilon^{-kt} \quad (1)$$

or the increasing exponential function (Fig. 2)

$$f(t) = 1 - \epsilon^{-kt} \quad (2)$$

The numeric ϵ has the value 2.71828. Therefore, the change in magnitude of $(1 - 1/\epsilon)$ has the fractional value 0 632121 Thus, after a time lapse of one time constant, starting at zero time, the magnitude of the physical quantity will have changed 63 2%

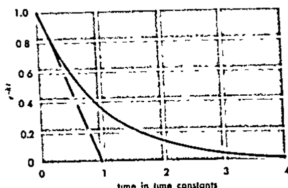


Fig 1 Universal time-constant curve for decreasing function

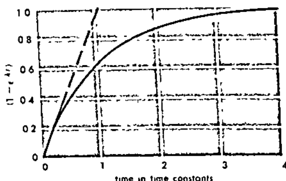


Fig 2 Universal time-constant curve for increasing function.

When time t is zero, Eq. (1) has the magnitude one, and when time t is $1/k$, the magnitude is ϵ^{-1}, or $1/\epsilon$. The corresponding change in magnitude is $(1 - 1/\epsilon)$. The specific time required to accomplish this change is

$$t = 1/k = T$$

T is called the time constant and is usually expressed in seconds. The same results are obtained for Eq. (2).

The initial rate of change of both the increasing and decreasing functions is equal to the maximum amplitude of the function divided by the time constant. Figures 1 and 2 are universal in that the plotted function is of unit height and the time scale is given in terms of time constants. To use these curves for a specific problem, the values in the ordinate axis are multiplied by the maximum amplitude of the quantity occurring in the problem, and the values in the abscissa axis are multiplied by the numerical value of the corresponding time constant

The concept of time constant is useful when evaluating the presence of transient phenomena. The relative amplitude of a transient after a lapsed time of a certain number of time constants is readily computed.

Lapsed time, time constants	*Transient completed, %*
1	63.2
2	86.5
3	95 0
4	98 2
5	99.3
10	99.996

Usually a transient can be considered as over after a period of 4–5 time constants

For electric circuits, the coefficient k and thus the time constant T is determined from the parameters of the circuit. For a circuit containing resistance R and capacitance C the time constant T is the product RC. When the circuit consists of inductance L and resistance R the time constant is L/R. *See* TRANSIENT, ELECTRIC.

The concept of time constant can be applied to the transient envelope of an ac signal; however it is more common to describe the change in amplitude in terms of the logarithmic decrement. For further discussion of this term, *see* DAMPING. [R L.R.]

Time-delay circuits

Electronic circuits producing an output signal delayed in time by a prescribed and controllable amount in relation to an input or controlling signal. One form of delay circuit that will reproduce the input waveform is the transmission line or its lumped circuit approximation (*see* DELAY LINE). More common forms of delay circuit are initiated by a controlling signal and produce an output, not necessarily related to the input in size or shape, at a later time. Usually the input pulses are recurrent at a specific rate, and therefore the time-displaced output signals are recurrent at the same rate. Such delay circuits are usually designed as linear delay circuits in the sense that a linear variation of some controlling element produces a delay that is a linear variation in time delay.

Multivibrator delay circuit. The cathode-coupled or emitter-coupled monostable multivibrator shown in Fig. 1 may be used as an approximately linear delay circuit (*see* MULTIVIBRATOR). For a fixed value of RC product, the duration T of the output waveform at each plate is proportional to the value of R, which can be calibrated in terms of a desired scale factor. The basis for such control is the fact that the magnitude of current in the plate circuit of VT-1 when it is conducting is proportional to its grid voltage. In turn, the voltage drop at the time the input trigger is applied is proportional to this current, and the time required for the multivibrator to recover to its initial state is proportional to this drop. The combination of variables involved leads to a pulse width T that is proportional to the input dc voltage level V. Thus a pulse

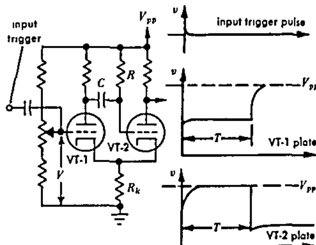


Fig 1. Monostable multivibrator as time-delay circuit.

front is generated at the output at a delay time T, with respect to the input pulse For sharp delayed trigger pulses, this output could be used to trigger a circuit, such as a blocking oscillator. *See* BLOCKING OSCILLATOR; TRIGGER CIRCUIT.

Although design considerations are somewhat different, two transistors in the common-emitter configuration may be used as a monostable multivibrator delay circuit in a manner similar to that employing vacuum tubes.

Linear sweep delay circuit. One of the most widely used forms of linear time-delay circuit makes use of a linear saw-tooth generator, such as the bootstrap or Miller integrator, whose output is then compared with a calibrated dc reference voltage level (*see* COINCIDENCE AMPLIFIER, COMPARATOR; SAW-TOOTH WAVE). If the saw-tooth voltage, which reaches a voltage equal to the desired reference voltage V_R in a time T, is applied to a comparator into which V_R is also an input, an output pulse will be obtained at a time T following the pulse that initiated the saw-tooth waveform

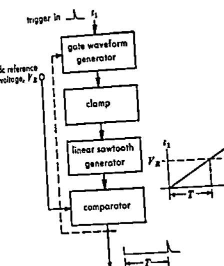


Fig. 2. Elements of linear sweep time-delay circuit

In some instances it is desirable that the entire

form generator and using the output pulse from the comparator to reset it to its initial state Elements of the time-delay circuit based on the use of the linear voltage sweep generator are shown in Fig. 2

Sanatron delay circuit. The basic sanatron delay circuit, of which there are a number of variations, combines in two pentode tubes the function of a ...te waveform generator, clamp and linear saw-...oth generator as shown in Fig. 3 The basic saw-...oth generator is the single-tube Miller integrator. T 2. Initially, its control grid is held at zero by ...id-current limiting No plate current flows, because the suppressor is held negative by the voltage-divider network from the plate of VT-1, which is heavily conducting The plate voltage is held at a definite starting level by the diode D-2 When a negative trigger pulse is applied to the suppressor grid of VT-1, its plate voltage rises, which in turn allows the suppressor grid of VT-2 to rise to a value, limited by the diode D-1, sufficient to allow plate current to flow in VT-2 This initial plate current drops the plate potential from the value set by

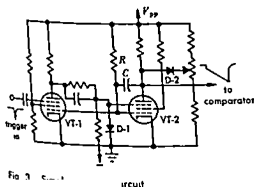


Fig. 3 S... ...ircuit

D-2, and causes a like drop at the grid. This drop at the grid, which is also connected to the grid of VT-1, maintains VT-1 nonconducting, thereby holding VT-2 in conduction. After the initial drop the tube VT-2 operates as a Miller integrator saw-tooth generator until some limit, dependent upon the operating levels of the tube, is reached. The saw-tooth output is then applied to a voltage comparator.

Phantastron. The phantastron combines the Miller integrator saw-tooth generator with the gating function. The output is applied to a comparator in a complete linear time-delay circuit as shown in Fig 4. Before the trigger is applied, the divider consisting of R_1, R_2, and R_3 holds the suppressor negative, which prevents plate current from flowing. All the space current, corresponding to zero grid voltage, goes to the screen. If a positive trigger pulse is applied to the suppressor, plate cur-

voltage sufficiently high to maintain plate-current

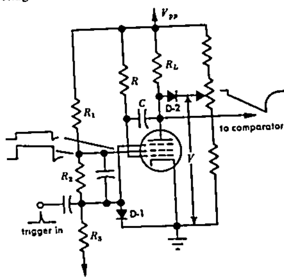


Fig 4 Screen-coupled phantastron delay circuit

flow The tube then functions as a normal Miller integrator The voltage level V from which the plate starts is determined by the divider to which diode D 2 is connected. This in turn determines the time required for the negative-going saw tooth to reach a given level When the plate voltage reaches saturation level, the grid voltage rises, and the circuit returns to its initial state.

A slightly different version of the phantastron is the cathode-coupled form, in which a cathode resistance provides the common coupling between the plate and screen circuits to allow switching to take place. *See* WAVE-SHAPING CIRCUITS. [G M G.]

Bibliography: B Chance, et al. (eds), *Electronic Time Measurements*, 1949; B. Chance, et al. (eds.), *Waveforms*, 1949, J Millman and H. Taub, *Pulse and Digital Circuits*, 1956.

Time-interval measurement

The measurement of elapsed periods of time. There are two classes of timers, one measuring the time of day, examples of which are clocks and watches (*see* TIME), and the other measuring time intervals, of which the stopwatch is an example.

A large number of devices and methods are available for measuring time intervals, both short and long. These devices are based on the following principles:

1. Timers controlled by the acceleration of gravity include pendulum and water clocks and the ancient hour-glass. In addition to their use in clocks, pendulums have been used to time events in laboratories. *See* CLOCK.

2. Mechanical vibrations depending upon the constancy of the elastic properties of materials include vibrating reeds, tuning forks, quartz crystals, and the balance wheel-escapement units used in ordinary watches, clocks and stopwatches. Electric circuits are required to obtain time measurements with tuning forks and quartz crystals. Clocks based on quartz crystals are called crystal clocks.

3. Electrical oscillations depending upon the constancy of the circuit elements include the rate of discharge of a capacitance and the resonant oscillations in resistance-capacitance and inductance-capacitance circuits These are used in laboratories to measure short time intervals

4. The vibration of atoms as applied in the atomic clock measures time of day precisely. *See* ATOMIC CLOCK; MASER

5. The rotations of the members of the solar system about their axes and about the sun are used in measuring longer periods The rotation of the earth about its axis defines the length of the day; its rotation about the sun defines the year. Instruments for direct measurement include transit telescopes and sundials.

6 The velocity of light or of other electromagnetic radiation is used in time-interval measurement. Radio waves, which have the same velocity as light, may traverse cavity resonators, that is a fixed distance, and if the traverse is repetitive, result in an electrical output of a definite frequency, which can be used to measure small time intervals. The difference in the velocity of light and of electricity flowing in wires, utilizing the Kerr cell as a valve, is used for measuring very small time intervals The difference in time for an electric current to flow in two wires of different lengths can also be used

7. Radioactive decay can be used to measure both long and short intervals of geological time. When the process which maintains the equilibrium of radioactive carbon (carbon-14) in a material such as bones or trees with carbon-12 in the atmosphere is interrupted, the radioactive carbon gradually transforms at a known rate to a nonradioactive form. Measurement of the residual radioactive carbon is a measure of the time interval since equilibrium, now measurable to nearly 70,000 years. For estimating much longer geological time intervals, other methods involving radioactivity are applied; for example, it is possible to measure the amount of helium trapped in rocks containing known amounts of radioactive materials which produce the helium. Computation of the decay can be converted to a time interval. *See* GEOCHRONOMETRY.

8. The measurable rate of mechanical rotation of a body is used to measure time intervals. The rotating body can be made to give a periodic signal by a rotating mirror, for example; the reflection of light focused upon the mirror at one point of its rotation will give this signal. *See* CHRONOGRAPH; CHRONOMETER; CHRONOSCOPE. [W.C.B.]

Timothy

A plant, *Phleum pratense*, of the order Graminales, long the premier hay grass for the cooler temperate humid regions. It is easily established and managed, produces seed abundantly, and grows well in mixtures with alfalfa and clover (*see* ALFALFA; CLOVER). It is a short-lived perennial, makes a loose sod, has moderately leafy stems 2–3 ft tall



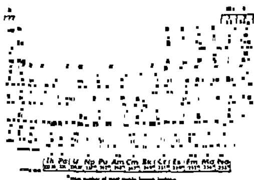
Timothy, *Phleum pratense*

and a dense cylindrical inflorescence (*see* INFLORESCENCE, PERENNIAL PLANTS). Timothy responds to fertile soils in yield and nutritive content. Cutting promptly after heading improves timothy's feed quality. Timothy-legume mixtures are the standard

hay members *of crop* rotations for the northern half of the United States. Timothy is also useful in pasture rotations. *See* GRAMINALES; GRASS CROPS. [H B S.]

Tin

Chemical element number 50, tin, Sn, is a member of group IV of the periodic table. It forms tin (II) or stannous (Sn^{2+}), and tin(IV) or stannic (Sn^{4+}), compounds, as well as complex salts of the stannite (M_2SnX_4) and *stannate* (M_2SnX_6) types.



Evidence of the earliest use of tin by man dates back over 4000 years The ancients found that tin has many unique properties other metals do not have. They were quick to realize that it alloys readily with copper to produce bronze It melts at a low temperature, is highly fluid when molten, and has a high *boiling point. It is* soft and pliable, and does not corrode.

The most important use is for tinplating steel containers for preserving foods. The next important uses are in solder alloys, babbitt (bearing metal), bronzes, brasses, type metals, and pewter Tin chemicals and compounds, both inorganic and organic, find extensive use in the electroplating, ceramic, and plastics industries. *See* TIN ALLOYS

Natural occurrence. The important tin-producing countries are Malaya, Indonesia, Bolivia, Thailand, Belgian Congo, Nigeria, and China.

Only one tin-bearing mineral, cassiterite (SnO_2), is of commercial importance. There are no high-grade tin ores. The bulk of the world's tin ore is obtained from low-grade alluvial deposits averaging about ½ lb of cassiterite per cubic yard (3000 lb) *Lode deposits* containing up to 4% tin are confined to Bolivia and Cornwall, where the cassiterite is associated with granitic rock and complex sulfides. *See* CASSITERITE.

Mining and concentration. The cassiterite is recovered from alluvial deposits by dredging in a placer, water jets and gravel pumps on level ground, or hydraulicking where a head of water permits it, and open-pit mining. The fine grains of cassiterite have a density 2½ times that of the gravel, and concentration is a simple matter *of screening and* gravity separations The concentrates contain 70–77% tin.

Underground lode deposits in Bolivia are located 12,000 ft above sea level. Access to these lodes follows the usual pattern of shaft sinking and driving adits. The ore is broken from the working face by drilling and blasting, and waste rock *is* disposed of below ground. Complex ore dressing methods and high transportation costs make Bolivia *the highest cost producer. See* MINING, PLACER; MINING, UNDERGROUND; ORE DRESSING; TIN METALLURGY.

Properties. Two allotropic forms exist: white tin (β) and gray tin (α). Although the transformation temperature is 13.2°C, the change does not take place unless the metal is of high purity, and only when the exposure temperature is well below 0°C. Commercial grades of tin (99.8%) resist transformation because of the inhibiting effect of the small amounts of bismuth, antimony, lead, and silver present as impurities.

Tin reacts with both strong acids and strong *bases, but it is relatively resistant to solutions that* are nearly neutral In the absence of oxygen, the high overpotential of tin causes a film of hydrogen to be retained on the surface so that attack by acids is retarded. A thin film of stannic oxide forms on exposure to air and provides surface protection

Halogen acids attack tin, particularly when they are hot and concentrated Hot sulfuric acid dissolves the metal, especially in the presence of oxidizing agents. Nitric acid attacks tin slowly when cold and dilute, more rapidly with rising temperature and concentration.

Dilute solutions of ammonium hydroxide and sodium carbonate have little effect on tin, but a strong alkali, such as sodium hydroxide, dissolves tin to form a stannate.

Salts that have an acid reaction in solution, such as aluminum chloride and ferric chloride, attack tin in the presence of oxidizers or air. Nonaqueous mediums have little effect on tin.

Properties of tin

Melting point, °C	231.9
Boiling point, °C	2270
Specific gravity, α-form (gray tin)	5.77
β-form (white tin)	7.29
Liquid at melting point	6.97
Transformation temperature, °C	13.2
Specific heat, cal/g, white tin at 25°C	0.053
Gray tin at 10°C	0.049
Latent heat of fusion, cal/g	14.2
Latent heat of vaporization, cal/g	520 ± 20
Heat of transformation, cal/g	4.2
Thermal conductivity, cal/(cm)(cm²)(°C)(sec), white tin at 0°C	0.150
Coefficient of linear expansion, at 0°C	19.9×10^{-6}
Shrinkage on solidification, %	2.8
Resistivity of white tin, microhms/cm³,	
at 0°C	11.0
at 100°C	15.5
Brinell hardness, 10 kg/(5mm)(180 sec),	
at 20°C	3.9
at 220°C	0.7
Tensile strength as cast, psi, at 15°C	2100
at 200°C	650
at −40°C	2900
at −120°C	12,700

Tin metal and the simple inorganic salts of tin are nontoxic. Tests have shown that concentrations of tin well above allowed limits in canned goods can be consumed without adverse effects on the human system. Some forms of organotin compounds, on the other hand, are toxic. The most important constants for tin are shown in the table.

Uses and applications. The United States is by far the largest consumer of tin. Coatings, alloys, and compounds are the most important outlets for tin.

Pure tin and alloys of tin can be applied as coatings to all the common metals by hot dipping or by electrodeposition. The tin gives protection to metal surfaces that oxidize or corrode readily. The coating also aids in fabricating and joining metals, and provides a clean adherent base for paint or lacquers.

Tin-coated steel (tinplate for cans) accounted for 1 ton out of every 20 tons of steel produced in the United States in 1957. More than 5,000,000 tons of tinplate and 33,000 long tons of tin were needed to make the 46,000,000,000 tinned-steel containers used by industry in 1957 in the United States, of which 60% were food cans and 40% were nonfood cans.

Tinplate manufacture is now largely a continuous electrolytic process with only a small percentage of production in hot tinning machines. Hot-dip tinplate normally carries a tin coating of 0.0001 in., and single sheets of steel are fed through the machine. The electrolytic tinning process is capable of handling continuous strands of strip steel at high speeds. Electrolytic tinplate can be produced in any desired coating thickness from 0.0001 to 0.000015 in. on each side. A thick coating on one side and a thin coating on the other is also possible (differential plate). The electrolytic process uses less tin per ton of steel, but tinplate consumption has increased annually by almost 3,000,000 tons since the first electrolytic line went into production in 1941.

Hot-dip tinplate amounted to only 13% of the total production in 1957. It is still used for special corrosive food packs. Heavily coated tinplate is also used for kitchen utensils, gas-meter cases, automotive parts, and returnable containers. Electrolytic tinplate, now 87% of production, finds wide use for general line food containers and for packaging nonfood products. Electrolytic tinplate in the thinner coating weights usually requires a baked enamel coating over the tin, except when used for mildly corrosive and dry food packs and for nonfood products. See FOOD PRESERVATION.

For other industrial applications, hot-dip tin coatings are applied to copper wire and sheet and to steel and cast iron parts. Examples are lugs and connectors for the electrical industry, the 10-gal tinned steel milk can, and tinned cast-iron food grinders. Hot-dip tin-lead coatings find service as a coating for gasoline tanks and filler pipes, and capacitor and transformer cans.

The plating industry consumes tons of tin as anodes for the electrodeposition of pure tin coatings and tin alloy coatings. Electrodeposited tin coatings, which can now be plated in a bright condition, make steel, copper, and aluminum easy to solder. Tin alloy coatings (tin-copper, tin-lead, tin-zinc, tin-cadmium, and tin-nickel) have advantages over single metal plates. They are denser and harder, more corrosion resistant, brighter or more easily buffed, and more protective for the basis metal. Tin-copper coatings (12% tin) have an appearance of 24-k gold, and when lacquered, serve as an attractive finish for jewelry, handbag frames, wire goods, and hardware. Tin-lead electroplates (10–60% tin) have excellent corrosion resistance and solderability, and are well adapted to the plating of printed circuits and electronic parts. Tin-zinc coatings (75% tin) have wide applications for radio, television, and electronics. They provide galvanic protection to steel in contact with aluminum. Tin-cadmium coatings (25% tin) are especially resistant to salt vapors, and have a number of applications in the aircraft industry. A tin-nickel coating (66% tin) will substitute for nickel and for copper-nickel-chromium as an ornamental finish. Commercial applications include coatings for watch parts, surgical and scientific instruments, coffee percolators, and costume jewelry. See METAL COATINGS.

Principal compounds (inorganic). Stannous and stannic tin salts have a number of uses in the field of electroplating, ceramics, and textiles. Those that are produced on a commercial scale are stannous compounds, oxides, and stannates.

Stannous oxide, SnO, forms black crystals soluble in acids and strong alkalies. It is prepared by treating stannous chloride with alkali. The precipitated stannous hydroxide is converted to the oxide by heating near the boiling point of water at a controlled pH. It is thermally stable up to 385°C, at which temperature, it is converted to stannic oxide. It is used in making stannous salts for plating and glass manufacture.

Stannic oxide, SnO_2, is a white powder, insoluble in acids and alkalies. It is prepared by atomizing tin with high-pressure steam and burning the finely divided metal, or by calcination of the hydrated oxide. It is an excellent glaze opacifier, a component of pink, yellow, and maroon ceramic stains and of dielectric and refractory bodies. It is an important polishing agent for marble and decorative stones.

Stannous chloride, $SnCl_2$, is available in the anhydrous and hydrated forms. The anhydrous salt results from the direct reaction of chlorine and molten tin. The hydrated salt is prepared by treating flaked tin with hydrochloric acid, followed by evaporation and crystallization. It is the major ingredient in the acid electrotin plating solution, and is an intermediate for tin chemicals.

Stannic chloride, $SnCl_4$, a fuming liquid, is prepared by direct chlorination of tin. The pentahy-

drate is a white solid. It is used in the preparation of organotin compounds and chemicals to weight silk and stabilize perfume and colors in soap.

Sodium stannate, $Na_2SnO_3 \cdot 3H_2O$, and potassium stannate, $K_2SnO_3 \cdot 3H_2O$, are prepared by dissolving hydrated stannic oxide in alkali. The sodium salt is a by product of the detinning of tinplate scrap. The stannates are used in alkaline electrotinning baths.

Heavy metal stannates of lead, barium, calcium, and copper are important in the manufacture of capacitor bodies.

Stannous sulfate, $SnSO_4$, is used in lacquer finishing steel wire and in electroplating.

Stannous fluoride, SnF_2, a white water-soluble compound, is a toothpaste additive.

Principal compounds (organic). Organotin compounds are those compounds in which tin is linked directly to one or more carbon atoms. Industrial interest started with the discovery that certain dibutyltin compounds, notably the dilaurate and maleate, were effective in preventing the decomposition of polyvinyl chloride resins during processing. Incorporating 2% of the stabilizer was found to be satisfactory, and several hundred tons of dibutyltin compounds are used annually in resins and chlorinated rubber paints.

The starting material for the commercial preparation of tin stabilizers is dibutyltin dichloride, which can be hydrolyzed to dibutyltin oxide. Quantitative yields are obtained by treating the oxide with equivalent amounts of the appropriate acids. In a modified Wurtz reaction, the dibutyltin dichloride, butyl chloride, and sodium react to form tetrabutyltin. This in turn is treated with stannic chloride to give dibutyltin dichloride, part of which is recycled. A number of diaryl and dialkyl compounds of tin are produced commercially.

Certain ranges of trialkyltin and triaryltin compounds possess powerful biocidal properties. These properties are possessed in high degree only when the tin atom is combined directly with three carbon atoms, as in the trialkyl compounds, they are at a maximum when the total number of carbon atoms in the molecule is about 12. The practical use of these compounds is in the fields of fungicides, insecticides, and pest control in growing crops. Tributyltin acetate $(C_4H_9)_3Sn{-}OOCCH_3$ and bistri-n-butyltin oxide $(C_4H_9)_3Sn{-}O{-}Sn(C_4H_9)_3$ are commercially available for use as antimicrobial agents in the fields of paper, wood, plastics, leather, and textiles. These compounds can be made by the modified Wurtz reaction and subsequent hydrolysis

compounds. Organotin hydrides (R_3SnH), prepared by the action of lithium aluminum hydride on the appropriate halide, enter into reactions with olefins to form tin-carbon bonds. When functionally substituted olefins are used, functionally substituted organic groups can easily be attached to the tin atom at moderate temperatures with yields of 80–90%. A large number of compounds are under study for an evaluation of their biocidal activity and industrial use. See ORGANOMETALLIC COMPOUND.

of 99.75% tin, and the user rarely needs to know the amounts of the different impurities.

The most satisfactory gravimetric method for tin is based on the precipitation of a tin-tannin complex from solutions of controlled acidity and subsequent ignition of the complex to SnO_2.

Volumetric methods utilize the reducing power of stannous compounds. Tin and tin alloys are dissolved in hydrochloric acid, the tin is reduced by digesting with nickel, granulated lead, or iron powder, and the solution is cooled in a protecting atmosphere of carbon dioxide and titrated with standard iodine or iodate solution. See ALLOY; BRAZING; SOLDERING. [R.M.M.]

Bibliography: W. E. Hoare, *Tinplate Handbook*, Tin Research Inst. Publ. 181, 1957; G. J. M. van der Kerk and J. G. A. Luijten, *Investigations on Organo-Tin Compounds*, Tin Research Inst. Publ. 221, 1957; J. W. Price and W. C. Coppins, *Sampling and Analysis of Tin Ingots*, Tin Research Inst. Publ. 195, 1955.

Tin alloys

Alloys account for about one-half of the world's consumption of new tin. These cover a wide composition range and many applications because tin forms alloys readily with nearly all metals (see TIN).

Soft solders constitute one of the most widely used and indispensable series of tin-containing alloys. Common solder is an alloy of tin and lead, usually containing 20–70% tin (see SOLDERING). It is made easily by melting the two metals together. With 63% tin, a eutectic alloy melting sharply at 361°F is formed. This is much used in the electrical industry. A more general purpose solder, containing equal parts tin and lead, has a melting range of 56°F. With less tin, the melting range is increased further and wiping joints, such as plumbers make, can be produced. Lead-free solders for special uses include tin containing up to 5% of either silver or antimony for use at temperatures somewhat higher than tin-lead solders, and tin-zinc base solders often used in soldering aluminum.

Bronzes are among the most ancient of alloys and still form an important group of structural metals. Of the true copper-tin bronzes, up to 10% tin is used in wrought phosphor bronzes, and from

5 to 10% tin in the most common leaded-tin cast bronzes. Many brasses, which are basically copper-zinc alloys, contain 0.75–1.0% tin for additional corrosion resistance in such wrought alloys as Admiralty Metal and Naval brass, and up to 4% tin in cast leaded brasses (*see* COPPER ALLOYS). Among special cast bronzes are bell metal, historically 20–24% tin for best tonal quality, and speculum, a white bronze containing 33% tin that gained fame for high reflectivity before glass mirrors were invented.

Babbitt or bearing metal for forming or lining a sleeve bearing has been one of the most useful tin alloys. It is tin containing 4–8% each of copper and antimony to give compressive strength and a structure desired for good bearing properties. An advantage of this alloy is the ease with which castings can be made or bearing shells relined with simple equipment and under emergency conditions. *See* BEARING, ANTIFRICTION.

Pewter is an easily formed tin-base alloy that originally contained considerable lead. Thus, because Colonial pewter darkened and because of potential toxicity effects, its use was discouraged Modern pewter is lead-free. The most favorable composition, Britannia Metal, contains about 7% antimony and 2% copper. This has desired hardness and luster retention, yet it can be readily cast, spun, and hammered.

Type metals are lead-base alloys containing 3–15% tin and a somewhat larger proportion of antimony. As with most tin-bearing alloys, these are used and remelted repeatedly with little loss of constituents. Tin adds fluidity, reduces brittleness, and gives a structure that reproduces fine detail.

Among miscellaneous tin-containing alloys commonly encountered are: costume jewelry, alloys similar to pewter and bearing-metal compositions which are often cast in rubber molds; die castings of tin hardened with antimony and copper for applications requiring close tolerances, thin walls, and bearing or nontoxic properties; and low-melting alloys for safety appliances The most common dental amalgam for filling teeth contains 12% tin. *See* ALLOY. [B.W.C.]

Tin metallurgy

Tin is comparatively easy to reduce to metal from the oxide by pyrometallurgy. However, this operation differs from the smelting of most common metals because it requires retreatment of the slag to obtain efficient metal recovery.

Smelting of tin concentrates is usually done in reverberatory furnaces using coke or coal as the reducing agent. The older method of using small shaft or blast furnaces is still practiced in a few localities, such as in the Far East. Also, electric furnaces have been used successfully, particularly in the Belgian Congo, but these account for only a small part of total production. The chief smelters are located at Penang and Singapore in the Malay Peninsula, Liverpool in England, Hoboken in Belgium, Manono in the Belgian Congo, and Texas City in the United States.

From alluvial tin ores, very high-grade cassiterite (natural tin oxide) concentrates are normally obtained which con[illegible] metal impurities. [illegible] reverberatory furn[illegible] vein deposits, such as those in Bolivia, are lower grade and usually contain substantial amounts of harmful impurities. If the grade is low, and Bolivian concentrates have ranged from 18 to 60% tin, additional upgrading at the smelter may be necessary. Sulfur, arsenic, and some lead, antimony, and bismuth are removed by roasting; addition of salt to form volatile or soluble chlorides assists particularly in removing lead and silver. Excessive amounts of iron and copper, with other soluble impurities, are leached from the calcine with hydrochloric (muriatic) acid. Various combinations of treatment fit specific needs.

In primary smelting, at about 1200–1300°C to keep the slag fluid, the amount of reducing agent is limited to give incomplete reduction and, thus, to produce metallic tin, low in iron. The resultant rich slag is consequently high in tin and usually high in iron. By a strongly reducing retreatment, with addition of more iron if necessary, a low tin slag is secured. This is discarded or retreated to recover prills (nuggets) of tin-iron. Because considerable iron is reduced in slag retreatment, the iron-containing tin, or hardhead, is fed back to the primary smelting furnace directly; or, after liquation to remove part of the tin, the high-iron dross may be recirculated to the slag-treatment furnace

Crude tin from smelting is liquated (that is, partially melted) to remove iron, copper, and other impurities which form solid compounds appreciably above the melting point of tin. At times, liquid tin has been filtered to reduce the iron content to an insignificant level Final refining is done in poling kettles by agitating the molten tin with steam or compressed air or with poles of green wood. Some metal impurities with considerable tin and tin oxide form a scum which is removed and recirculated through the smelting cycle The refined metal from most smelters is over 99.8% tin Recovery of tin is usually over 97%, and discarded slag contains only 1–2% tin

Secondary tin from metal scrap amounts to about one-third of the total tin consumed in the United States Most of it comes from tin-bearing alloys, and secondary smelters rework them into alloys and chemicals However, several thousand tons per year of tin metal of highest purity is recovered from the detinning of tin-plate scrap The tin is removed with a hot caustic solution and recovered electrolytically. *See* PYROMETALLURGY, NONFERROUS, TIN [B.W.C.]

Bibliography: C. L. Mantell, *Tin*, 2d ed., 1949.

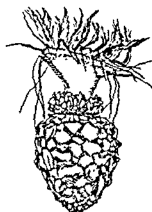
Tinamiformes

An order of birds comprising the single family Tinamidae, the tinamous of Central and South

[A]merica. Nine genera, containing about 35 species, [a]re recognized. Tinamous are superficially fowl-like [bi]rds of uncertain affinities. There is skeletal evi[d]ence to suggest that they share a common ances[to]r with the rheas (Rheiformes), and it is interest[in]g to note that in tinamous, as in rheas and most [o]ther ratites, the males incubate the eggs and care [fo]r the young. Although they have fully developed [wi]ngs, tinamous are weak flyers and prefer to run [o]r hide to escape danger. Tinamou eggs are nota[b]le for their glossy, often deeply colored, shells. [S]ee AVES. [K.C.P.]

Tintinnida

An order of the Spirotricha which are conical or trumpet shaped pelagic forms that live in shells or [l]oricae. They are especially abundant in oceans such as the Pacific. The exact structure, often quite elaborate, and the dimensions of the lorica are so recognizably different among the hundreds of known genera that the taxonomic arrangement of forms within the order is based solely upon characteristics of this secreted "house." The adoral

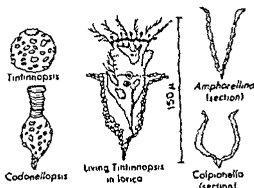


Tintinnopsis, an example of a tintinnid.

zone of membranelles is prominent, while the other ciliature is greatly reduced. Fossilized tintinnids, actually only the loricae have been preserved, represent practically the only fossil ciliates known to science. *Tintinnus*, *Tintinnopsis* (see illustration) are common. See SPIROTRICHA; TINTINNINA. [J.O.C.]

Tintinnina

A suborder of planktonic marine ciliated protozoans of the class Ciliata, characterized by a trumpet shaped body in a secreted outer shell, or lorica. The lorica is composed of a resistant organic compound in which are embedded various foreign mineral grains; is commonly trumpet- or bell-shaped, cylindrical, or subspherical, and ranges in ... ms. *Tintinnopsis* is the



Fossil and modern Tintinnina: *Tintinnopsis*, Jurassic to Recent; *Codonellopsis*, Recent; *Amphorellina*, Lower Cretaceous; *Colpionella*, Recent.

most c ... identi ... cross ... tions of the rocks in which they are found. Twelve genera of fossil tintinnids have been described from limestones and cherts of the Jurassic and Cretaceous. See CILIOPHORA. [D.J.J.]

Bibliography: A. S. Campbell and R. C. Moore, *Treatise on Invertebrate Paleontology*, Part D, Protista 3 (Radiolarians, Tintinnines), 1955.

Tire

The separate circumferential portion of a wheel, designed to roll in contact with the surface over which the wheel travels. The principal type of tire in use today on automotive vehicles is the pneumatic tire, comprising an outer case or shoe, either sealed to the rim of the wheel to contain air under pressure, or fitted with a separate inner tube that contains the air. Such tires are used on road vehicles such as cars and trucks, on off-the-road vehicles such as earth movers, and on airplanes and bicycles. The tire cushions the vehicle (and the road or runway) from shock, provides traction, resistance to skidding, and cornering power, whereby the tire transmits a lateral force between vehicle and surface to enable the vehicle to turn curves. [N.M.]

Tissue

An aggregation of cells more or less similar morphologically and functionally. The animal body is composed of four primary tissues, namely epithelium, connective tissue (including bone, cartilage, and blood), muscle, and nervous tissue. The process of differentiation and maturation of tissues is called histogenesis. See HISTOLOGY; PLANT TISSUE SYSTEMS. [C.B.C.]

Titanate

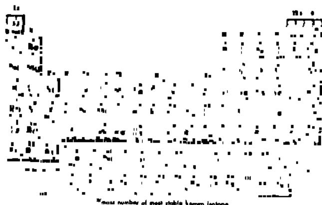
A compound obtained when metal oxides or hydroxides are heated with titanium dioxide, TiO_2. Metatitanates of the formulas K_2TiO_3, $ZnTiO_3$, $PbTiO_3$, and $BaTiO_3$ are formed by fusion of TiO_2

with the appropriate metal oxide. Titanates of the Na_4TiO_4 type are often called orthotitanates.

The titanates are insoluble in water and have found use in the manufacture of porcelain enamels, ceramic dielectrics, temperature-sensitive resistors, and photoconductors. *See* CERAMIC TECHNOLOGY; TITANIUM. [E.E.WR.]

Titanium

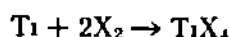
Chemical element number 22, titanium, Ti, occurs in the fourth group of the periodic table, and its chemistry shows many similarities to that of silicon and zirconium. On the other hand, as a first-row transition element, the aqueous solution chemistry,



*mass number of most stable known isotope

especially of the lower oxidation states, shows some resemblances to that of vanadium and chromium *See* TRANSITION ELEMENTS.

The outer electronic arrangement is $3d^24s^2$, and the principal valence state, correspondingly, is 4+; the 3+ and 2+ states are also known, but are less stable. The element burns in air, when heated, to give the dioxide, TiO_2, and combines with halogens according to the reaction



$$Ti + 2X_2 \rightarrow TiX_4$$

It reduces water vapor to form the dioxide and hydrogen, and reacts similarly with hot concentrated acids, although with hydrochloric acid, the trichloride is formed The metal absorbs hydrogen to give compositions approaching TiH_2 and forms the nitride, TiN, and the carbide, TiC. The sulfide TiS_2 can be formed by the reaction of $TiCl_4$ with H_2S, and both the lower oxides, Ti_2O_3 and TiO, and sulfides, Ti_2S_3 and TiS, are known.

Salts of all three valence states are known The yellowish $Ti(SO_4)_2$ is a prominent compound, but except in very highly acid solutions, the titanium hydrolyzes to titanyl ion, TiO^{2+}. There is, for example, a basic phosphate, $KTiOPO_4$ With respect to the lower valence states, the halide salts are the best known and give Ti^{2+} and Ti^{3+} (titanous) ions *in solution*. The latter is violet in color with a broad spectral absorption band at 490 mμ; this has been interpreted as being caused by a splitting of the 3*d* orbitals by the electrostatic field of the 6 water molecules which surround the ion.

Natural occurrence. The dioxide, TiO_2, occurs most commonly in a black or brown tetragonal form known as rutile (a/b ratio = 4.58/2.95 A). Less prominent naturally occurring forms a anatase (also tetragonal, a/b = 3.73/9.37 A) a brookite (rhombohedral). Both rutile and anata are white when pure. The dioxide may be fus with other metal oxides to yield titanates, for e ample, K_2TiO_3, $ZnTiO_3$, $PbTiO_3$, and BaTiC The black basic oxide, $FeTiO_3$, occurs naturally the mineral ilmenite; this is a principal commerci source of titanium. The dioxide, for example, m be prepared from it by dissolving the ore in sulfur acid, clarifying, and then partially neutralizing th solution. The titanium hydrogel which forms is the separated from the solution, compacted, dried, an calcined. Depending on the heat treatment, the r sulting powdered TiO_2 may be of the anatase (lo temperature) or the rutile (high-temperature type. The dioxide is refractory, and as rutile, ha the unusually high dielectric constant of about 10 (173 parallel to, and 89 perpendicular to, th principal axis) and an index of refraction of 2. (as compared to 2 42 for diamond) The anhydrou dioxide, although somewhat acidic, as evidenced b the formation of titanates, is quite inert to acids although the freshly prepared hydrous material i soluble in acid *See* ILMENITE

Principal compounds. The sesquioxide, Ti_2O_3 may be prepared by the hydrogen reduction of the dioxide it is acid soluble, to give solutions contain ing titanous ion, from which the black $Ti(OH)_3$ may be precipitated by addition of base. In the presence of water, $Ti(OH)_3$ evolves hydrogen to give the dioxide. Titanous ion itself is a good reducing agent. It has found use in volumetric analysis as a quantitative reagent for the determination of ferric and permanganate ions

The 2+ oxide, TiO, can be obtained by the high-temperature reduction of the dioxide by carbon or various metals It is basic, but its salts are unstable in water solution because of the strong reducing power of the Ti^{2+} ion.

Among the halogen compounds of titanium, the best known are the tetrahalides, TiX_4. In addition, the complex ions TiF_6^{2-} and $TiCl_6^{2-}$ are well known Titanium tetrachloride is a light yellow liquid boiling at 136°C; it may be prepared by the direct reaction of the elements, but a recent commercial process makes use of the common ore ilmenite as starting material. As in the manufacture of the dioxide, the ore is dissolved in sulfuric acid, but the solid K_2TiCl_5 is precipitated out by saturating the solution with HCl and KCl. The complex salt is then thermally decomposed to give $TiCl_4$ The tetrachloride hydrolyzes with water or moist air to give the dioxide, it reacts with metal trialkyls, such as aluminum triethyl, to form $TiCl_3$ and more complex compounds, and with alcohols to form compounds of the type $Ti(OR)_4$

The trichloride may also be obtained by reduction of $TiCl_4$ by metals such as silver or zinc, and by electrolysis. The dichloride can be prepared by the thermal decomposition of the trichloride.

Titanium in the 4+ state forms various complex ions, in addition to the TiX_6^{2-} species described

above. Chelate compounds with cupferron, 8-hydroxyquinoline, acetylacetone and its derivatives, and quinone are known. Finally, the orange peroxy ion, $TiO_2(SO_4)_2^{2-}$, is formed by the addition of hydrogen peroxide to a solution of the sulfate, and ... acid H_4TiO_5 is known. The formation

...

paints because of its chemical inertness, superior covering power, opacity to damaging ultraviolet light, and self-cleaning ability. Both the rutile and anatase forms are used, but especially the former. The total production of TiO_2 in the United States amounts to some 400,000 tons annually, mostly from ilmenite ores.

The dioxide has also been used as a whitening or opacifying agent in numerous situations. Examples would be the use as a filler in paper, a coloring agent for rubber and leather products, a pigment in ink, and a component of ceramics. Recently, it has found important use as an opacifying agent in porcelain enamels, giving a finish coat of great brilliance, hardness, and acid resistance. Rutile has also been found as brilliant, diamondlike crystals, ... form

...

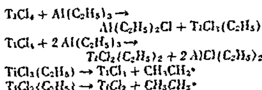
13 for $MgTiO_3$ to several thousand for solid solutions of $SrTiO_3$ in $BaTiO_3$. Barium titanate itself has a dielectric constant of 10,000 near 120°C, its Curie point, it has a low dielectric hysteresis. These properties are associated with a stable polarized state of the material analogous to the magnetic condition of a permanent magnet, and such substances are known as ferroelectrics. In addition to the ability to retain a charged condition, barium titanate is piezoelectric and may be used as a transducer for the interconversion of sound and electrical energy. Ceramic transducers containing barium titanate compare favorably with Rochelle salt and quartz, with respect to thermal stability in the first case, and with respect to the strength of the effect and the ability to form the ceramic in various shapes, in the second case. The compound has been used both as a generator for ultrasonic vibrations and as a sound detector. See PIEZOELECTRICITY.

Although somewhat corrosive, liquid titanium tetrachloride has found use in the formation of smokes, especially in World War I, and also in commercial skywriting. On contact with moist air, the compound $TiCl_4 \cdot 5H_2O$ first forms, followed by hydrolysis to the dioxide.

More recently, $TiCl_4$ has become an important starting material for the production of titanium metal (by means of magnesium or sodium reduction). Its commercial preparation from ilmenite has already been described. The compound has become very important in the catalytic polymerization of ethylene. ...

Titanium es... with alcohols, ... ful as waterproofing agents ... and synthetic fabrics. The tetrabutyl and tetraisopropyl esters hydrolyze in moist air to give the dioxide, and can be used to provide thin, transparent, and adherent coatings. The diacetate, $TiCl_2(O_2C_2H_3)_2$, has been suggested as a flame retardant for cellulose fabrics. The acetylacetonate, $Ti(C_5H_7O_2)_2$, may be used as a crosslinking agent in lacquers so that on drying, the resulting film becomes inert to solvents.

Polymerization catalysis. An important development in the low-pressure polymerization of ethylene has been the use of titanium catalysts. Typical starting materials would be $TiCl_4$ and $Al(C_2H_5)_3$, which then react according to the general scheme:


$$TiCl_4 + Al(C_2H_5)_3 \rightarrow Al(C_2H_5)_2Cl + TiCl_3(C_2H_5)$$
$$TiCl_4 + 2\,Al(C_2H_5)_3 \rightarrow TiCl_2(C_2H_5)_2 + 2\,AlCl(C_2H_5)_2$$
$$TiCl_3(C_2H_5) \rightarrow TiCl_3 + CH_3CH_2\cdot$$
$$TiCl_2(C_2H_5)_2 \rightarrow TiCl_2 + CH_3CH_2\cdot$$

At the same time, mixed halide-alkyl complexes of variable composition are formed, for example, $(C_2H_5)_2TiCl_2Al(C_2H_5)_2$. It is the catalytic activity of such complexes that forms the basis of the well-known Ziegler process for the polymerization of ethylene. This type of polymerization is of great industrial interest since, by means of it, high-molecular weight polymers can be formed. In some cases, desirable special properties can be obtained by forming isotactic polymers, or ones in which there is a uniform stereochemical relationship along the chain. See POLYOLEFIN RESINS; TITANATE; TITANIUM METALLURGY. [A.W.A.]

Titanium metallurgy

Titanium is the fourth most abundant metallic element in the earth's crust and ninth most common element. It is a silvery-gray, paramagnetic metal produced in strengths equal or superior to steel, although its specific gravity (4.5) is approximately 56% that of alloy steel. It retains its properties in the temperature range −320 to +1000°F. Its melting point is 3074°F; boiling point, 6395°F; coefficient of expansion, 8.5×10^{-6}. Titanium metal is low in electrical and thermal conductivity, and it also offers outstanding resistance to corrosion in oxidizing media and is impervious to atmospheric or salt-water corrosion.

The unique combination of titanium's intrinsic properties was immediately seized upon by designers seeking to lighten the weight of jet engines and airframes when wrought titanium shapes (sheet, strip, plate, bar, billet, wire, extruded shapes, and tubing) became available in 1950. Since then, titanium metal has achieved an industrial status that took other metals, such as lead, copper, and zinc, 40–80 years to reach. Of the newer metals, aluminum spanned 28 years and magnesium 26

years in arriving at the production capabilities offered by the titanium industry in 1959.

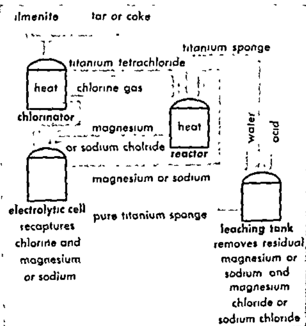
Titanium was first discovered in 1791, when an English clergyman and amateur chemist, William Gregor, separated it from the black magnetic sands of Cornwall.

Metallurgical extraction. Crude titanium was first isolated in 1825 by J. J. Berzelius, but it was not until 1906 that M. A. Hunter separated enough metal for study. Titanium metal (as contrasted to titanium pigment) as known today was first produced by a process patented by William A. Kroll, a Luxembourg scientist. Kroll first made metallic titanium in 1928 while looking for a substitute for beryllium for copper-beryllium alloys. He found no commercial interest in the new material in either Europe or the United States.

In 1937, Kroll invented the reduction process which bears his name (the dr[illegible] erature [illegible] esium). [illegible] 2 tons of sponge by the Kroll process, and in 1957, total production was 17,500 tons.

Problems encountered in producing titanium initially appeared to be insurmountable. The liquid metal seems to be a universal solvent, and either dissolves or is contaminated by every known refractory; the metal must be reduced (won) from its ore with extreme purity, because the contaminants generally destroy the desirable physical properties; furthermore, the metal, when molten, is so very active chemically—absorbing nitrogen or oxygen from the air quite rapidly—that all extractive and ingot-melting processes must be carried out either in vacuum or under the protection of an inert atmosphere of helium or argon

The first step in winning titanium metal from its ore is to chlorinate an oxide-carbon mixture to obtain titanium tetrachloride This first step has many counterparts in extractive chemistry because it is a useful method for obtaining volatile or soluble compounds of many refractory metals, which can be separated from other constituents by fairly simple means and then reduced. Thus, the titanium tetrachloride is treated with magnesium metal in a heat-resistant steel vessel at a red heat under an inert gas blanket. The products of this reaction are commercially pure, spongy titanium metal and magnesium chloride. The bulk of the magnesium chloride is drained out of the reaction chamber as a liquid and is electrolyzed to recapture chlorine gas and magnesium metal.



Sponge processing. *(After R. A. Malasick)*

The titanium sponge, with scrap and alloy elements added, is then pressed into electrodes and melted into a primary ingot in a water-cooled copper crucible. This ingot is then remelted in a consumable-arc vacuum furnace into the final ingot. By double-melting with consumable arcs and with a reduced-pressure furnace atmosphere, electrode contamination is avoided and the hydrogen content is held to a very low level.

Forging, rolling, and drawing of titanium from ingot to finished product present no peculiar problems not mastered by mills experienced in the handling of stainless and high alloyed steels. In general, titanium requires smaller reductions than stainless steel to minimize edge cracking. *See* METAL FORMING.

Volume and markets. Before 1957, 90% of titanium metal production was restricted to military purposes, and estimates show that 98% of production actually was channeled to defense outlets, primarily for jet engine compressor parts and airframe structural assemblies. Until 1957, titanium production showed uninterrupted growth to 17,500 sponge tons The strategic shift from manned aircraft to unmanned missiles was reflected in the production in 1958 of 4500 sponge tons. Sponge prices were reduced from $5 00 per pound in 1950 to $1.62 in 1958 The Composite Price Index of Titanium Metals Corporation of America (based on commercially pure sheet and strip, alloy bar and billet, plus applicable price extras) shows mill product prices were reduced from $15 25 per pound to $7.59 per pound.

Because its price compares favorably with that of other vacuum-melted metals, titanium is now winning increased acceptance in the basic chemical industry for valves, pumps, heat-transfer units, and liners; in the electronic industry, also because titanium absorbs gases and has a coefficient of expansion similar to ceramics; in missiles, in nuclear reactors; and in commercial airplane manufacture.

The cost of fabricating titanium into finished assemblies is comparable to that of stainless steels. In general, these fabrication costs amount to about four times the cost of the steel

Titanium-base alloys. In general, titanium alloys fall into three classes depending on the phase (or phases) present at room temperature.

Alpha alloys. Titanium of the highest possible purity, commensurate with large-scale production

tolerable cost, is called commercially pure titanium Such metal is certainly by strict interpretation an alloy of titanium and various interstitial elements such as carbon, oxygen, and nitrogen. Because these interstitials total only about 0.25% of the composition, there is ample metallurgical precedent for the designation *commercially pure*.

There is only one commercial substitutional α-alloy, Ti 5Al 2.5Sn, containing 5% aluminum (the only metallic α-stabilizer) and 2.5% tin

Generally speaking, α alloys (hexagonal close-packed structure) have highest strength and best oxidation resistance at high temperatures (600–1100°F) and the best weldability of all titanium alloys. However, without a β phase, these alloys have lower strength at room and moderately elevated temperatures. Furthermore, they respond only slightly, if indeed at all, to heat-treatment.

At present, commercially pure and Ti-5Al-2.5Sn are employed where weldability and moderate strength, at or below 1000°F, are required

Alpha-beta alloys. The α-β-alloys vary widely in composition and general characteristics. At one end of the alloy range are deep-hardening alloys such as Ti-2Fe-2Cr 2Mo and Ti-4Al-4Mn, which provide high strength at room and moderately elevated temperatures. At the other end are the lean α-β compositions, such as Ti-6Al-4V and Ti-7Al-4Mo These alloys are shallow hardening, but their comparatively high aluminum content gives them high strength and improved elevated temperature properties.

As a class, α-β-alloys have high strength, respond better to heat treatment, and are more formable than α-alloys. Their weldability is inferior, however

High strength α-β-alloys have found diversified uses, particularly for forgings, fasteners, and sheet applications.

Beta alloys. The only metastable β-alloy (body-centered-cubic structure) in commercial produc- [illegible] be solution-treated, formed in the soft condition, and then age-hardened to high strength. The alloy has somewhat higher density than most other commercial alloys and is not thermally stable above 700 F. Its ability to be cold-worked makes it the first high-strength titanium alloy that can be used for cold-headed bolts. See ALLOY; PYROMETALLURGY, NONFERROUS; TITANIUM. [T W.L.]

Titmouse

Any of several species of the widely distributed songbird family Paridae, which also contains the chickadees. The United States has three species of titmice, all in the genus *Baeolophus* (= *Parus*) By far the best known is the tufted titmouse, *B bicolor*, a common bird of the eastern deciduous forest. It is the only eastern gray bird that has a crest; it is whitish below, and the flanks are washed with rust. The tufted titmouse usually winters throughout its breeding range. It is as acrobatic as the chickadee in its feeding, and is also well known for its clear, whistling call of "peter, peter, peter." See CHICKADEE. [J.D.B.]



The tufted titmouse, *Baeolophus bicolor*; length to 6½ in (From E L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Titration

The process in which a solution containing a known concentration of a substance (the titrant) is added to another solution containing an unknown concentration of a second material (the analyte) that will react with the titrant The titrant is added until there is some indication that an amount of substance equivalent to the material of unknown concentration has been added If the stoichiometry or the exact ratio is known for the manner in which the substance in the titrant reacts with or is equivalent to the material of unknown concentration, it is possible to calculate the amount of material present in the unknown solution See VOLUMETRIC ANALYSIS.

In order to perform a titration, it is necessary to have (1) a solution of known concentration and hence chemicals of known purity to prepare these solutions, (2) some means of detecting the completion of the reaction, and (3) calibrated apparatus, including burets, pipets, and volumetric flasks.

Solutions of known concentration, called standard solutions, are prepared in two ways. The direct method involves weighing a sample or measuring a volume of a pure substance (primary standard) into a volumetric flask, dissolving it in a suitable solvent, usually water, and then diluting it carefully to volume in the flask. The concentration of the solution is then calculated from the weight of substance taken and the volume to which the solution was diluted. The indirect method is used when

a pure substance or a substance of known purity is not available for preparing the solution. In these cases, a solution of approximately the concentration desired is prepared, and its concentration is determined by titration against a weight of substance of known purity.

The reliability of a titration depends upon the accuracy with which volumes of solutions can be measured. This is usually less than the accuracy of weighing, and is greatly affected by the temperature of the solution, the cleanliness of the measuring equipment, and the certainty of the markings on calibrated equipment. For the most accurate work in titrations, these errors must be minimized.

Classification by chemical reaction. The types of chemical reactions that can be followed by titration methods fall into three general categories.

Acid-base or neutralization reactions. These applications involve the titration of an unknown base with a titrant containing a known concentration of an acid or, conversely, the titration of an unknown acid with a basic titrant. For example, to determine the amount of acetic acid in vinegar by titration, a known volume or weight of vinegar is dissolved in water and titrated with a standard solution of sodium hydroxide. From the volume (and concentration) of sodium hydroxide solution added when the indicator, phenolphthalein, just changes from colorless to a faint pink color, the weight of acetic acid present in the sample can be calculated. Because any other acidic substance present in the vinegar would also react with the titrant, this determination of acetic acid cannot be considered a specific determination unless it is known that acetic acid is the only acidic substance present. This nonspecificity is characteristic of all titration methods, but by careful control of experimental conditions or by the prior use of various separation techniques, titration methods can be used to perform an analysis for a single component in a fairly complicated mixture. See ACID AND BASE; NEUTRALIZATION.

Oxidation-reduction reactions. In these applications, an oxidizing agent is used as the titrant for substances that undergo a stoichiometric oxidation. Conversely, a solution of a reducing agent may be used as the titrant for oxidizing agents. However, because many reducing agents are oxidized by oxygen from the air, standard solutions of many reducing agents are not stable. For this reason, solutions of reducing agents must be standardized frequently. Consequently, standard solutions of oxidizing agents are more widely used for titrations in this category than standard solutions of reducing agents. See OXIDATION-REDUCTION.

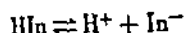
Precipitation and complex-formation reactions. Theoretically, all reactions that involve the formation of either a precipitate or a complex ion, if they proceed sufficiently rapidly and if the products are of definite composition, should provide the basis for a titrimetric determination. A precipitation reaction that is widely used for a titrimetric determination is the precipitation of chloride ion with silver ion from a standard solution of silver nitrate. There are various convenient and accurate methods to determine the equivalence point between these reactants. However, many precipitation reactions do not lend themselves to titration methods because the precipitate may form too slowly and is not of a definite and reproducible composition, and because there is no convenient method to detect the equivalence point. Titrations involving complex-ion formation have become much more important since the discovery of ethylenediaminetetraacetic acid and related compounds. Procedures now exist for the titration of more than 30 metal ions by using these reagents. Furthermore, by utilizing the techniques of masking (competitive complexation reactions) and pH control, it is possible to determine each of several metal ions in a mixture. See COMPLEX COMPOUNDS; PRECIPITATION (CHEMISTRY).

Solvents. Titrations need not be carried out in aqueous solutions. Considerable use has been made of nonaqueous media, principally for the titration of acids or bases. Substances that exhibit extremely weak basic properties in water will react as stronger bases in a more acidic solvent such as glacial acetic acid. Consequently, by using a titrant of perchloric acid dissolved in glacial acetic acid, it is possible to determine substances that are too weakly basic to react completely in aqueous media. Conversely, basic solvents such as ethylenediamine can be used for the medium of the titration of very weak acids. In addition, various ketones such as methyl isobutyl ketone, and even some aprotic liquids such as benzene have been used for the solvent in the titration of mixtures of weak acids or mixtures of strong and weak acids. In these solvents, the base is usually a tetraalkylammonium hydroxide.

Classification by end points. The accuracy of all titrations is limited by the ability to detect the equivalence point between the titrant and the analyte. The completion or end point of a titration is usually detected with the aid of some sort of indicator or by a physical measurement. Potentiometric methods can be considered the universal method for detecting the true equivalence point in a titration. From such measurements, the suitability of various chemical indicators, which are generally more convenient to use, can be deduced. Chemical indicators are chosen so their distinctive color change occurs as close as possible to the true equivalence point. The end point of the titration is then equal to the true equivalence point. If the end point does not correspond to the equivalence point, an error which is proportional to the discrepancy between the equivalence point and the point where the indicator changes color is introduced into the determination.

Indicators. Indicators used for acid-base titrations are usually weak organic acids or bases. These substances show sharp transitions from colored to colorless forms or from one colored form to another over definite and rather narrow changes in acid concentration. These color changes have been attributed to transformations from ion to molecule.

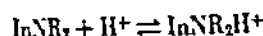
ch for an indicator that is a weak, monoprotic l can be represented:



$$HIn \rightleftharpoons H^+ + In^-$$

Acidic form Basic form

:nolphthalein is an example of this type of indi- or where the acidic form, the *undissociated* lecule is colorless and the basic form, the anion, red. Methyl orange is a weak organic base and tran-formation can be formulated:



$$InNR_2 + H^+ \rightleftharpoons InNR_2H^+$$

Basic form (yellow) Acidic form (red)

The color change of most acid-base *indicators is* read over about a 100-fold (2 pH units) change the hydrogen-ion concentration. Because the idrogen ion concentration at the equivalence point various acid base titrations may differ by as much s 10', it is obvious that the proper choice of in- icator is essential if the end point is to equal the quivalence point.

Chemical indicators are used also for oxidation- reduction, precipitation, and complexation titra- ions. Even though the chemical reactions involved with the use of these indicators are different from that outlined for acid base indicators the same precautions concerning the choice of the proper indicator must be observed in these cases.

Indicators usually provide the most convenient means for detecting *the end point of a titration* However, if the analyte is itself intensely colored the color change of the indicator cannot be ob- served. There are also a considerable number of titrations where there is no indicator that under- goes a sharp color change in the region of the equivalence point. In these cases, as well as where two or three titratable substances in the analyte may be titrated consecutively and each equivalence point is to be detected, a physical method instead of an indicator is used for following the course of

erties or in the heat content of the solution The chief electrical methods utilize the change in poten- tial *of* an indicating electrode or the change in electrical conductance during a titration *See* TI- TRATION, CONDUCTIMETRIC, TITRATION, POTENTIO- METRIC.

Optical methods for detecting the equivalence point of a titration depend *on observing the change* in the color or in the absorbance of the solution during the titration This may be done either visually or *instrumentally* If a titration is repre- sented by the general equation.

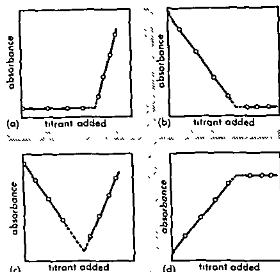


$$X + R \rightleftharpoons X' + R'$$

Analyte Titrant Products

and if the titrant is highly colored and the analyte and *products are* essentially colorless it should be evident that [illegible] nt color will occur in the

solution until after an amount of titrant equivalent *to the analyte has been added. An* example of this type of titration is the titration of iron(II) ions in dilute sulfuric acid solution with a standard solu- tion of potassium permanganate. Of course, if *visual detection* is used, a finite concentration of titrant in excess of the equivalence point must be added before sufficient color is present to be ob- served On the other hand, if the change of color is followed *instrumentally by measuring the amount of light* absorbed by the solution, the absorbance can be measured when various volumes of titrant have been added Prior to the equivalence point, the absorbance will be virtually zero, but after the equivalence point, the absorbance will usually in- crease linearly with the concentration of excess titrant added, as shown in part (*a*) of the illustra- tion; the point of *intersection of the two straight*



(a–d) End-point detection in photometric titration

lines will indicate the true equivalence point. Therefore this method of detecting the end point, which is referred to as a photometric titration, is more accurate than visual detection

In other titrations, the analyte may be colored, whereas the titrant and products are colorless. In such cases, the color of the solution would gradually decrease as the titrant is added and the photometric titration curve would be similar to that *shown in* part (*b*) of the illustration. If both titrant and analyte are colored but the products are colorless, the absorbance of the solution would decrease until the equivalence point is reached and would then in- crease as excess titrant is added. This type of titra- tion curve is shown in part (*c*) of the illustration. Finally, if the products of the titration absorb light, but the reactants do not, the absorbance will in- crease during the course of the titration until the equivalence point is reached and then either remain constant or decrease slightly because *of dilution.* This photometric curve is shown in part (*d*) of the illustration.

Even greater selectivity and accuracy can be achieved in photometric titrations if a spectrophotometer is used to measure the absorbance of the solution. The absorbance can then be measured at the wavelength of maximum absorption for one of the reactants or products in the titration and much less interference from the absorbance of other nonreacting substances is encountered. Furthermore, if these absorbance measurements are made with ultraviolet light, it is often possible to detect end points in titrations where there is little or no visible color change during a titration

As shown in the illustration, end points in photometric titrations are found by extrapolation. When the end point is detected visually either by the color change of an indicator or from the color of a titrant, it is necessary to titrate slowly in the vicinity of the end point so that the volume of titrant can be read when this color is just observed. Thus, this somewhat tedious process is avoided in photometric titrations and in any other technique for detecting end points which involves an extrapolation. *See* TITRATION, AMPEROMETRIC

Thermometric titrations. This process of detecting the end point of a titration by measuring the temperature of a solution as the titrant is added is called thermometric or enthalpy titrations. If a chemical reaction is exothermic, the temperature of the solution will rise during the titration until the equivalence point is reached. Further addition of titrant will cause relatively smaller changes in the temperature because no more reaction is occurring and any temperature change must result merely from the mixing of the titrant and the solution. For endothermic reactions, the temperature of the solution will fall until the equivalence point is reached.

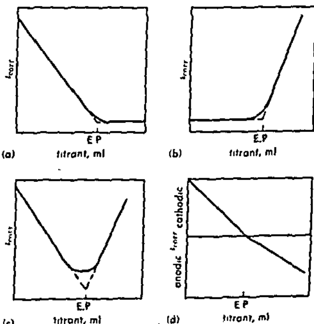
Because the heat evolved for dilute solutions is small, this technique cannot be applied to solutions more dilute than 0.002 *M*. Even with more concentrated solutions, the temperature change during a titration is small. However, if precautions are taken to avoid transfer of heat between the titration vessel and its surroundings and if the temperature of the titrant and the initial temperature of the analyte are equalized, end points can be detected by measuring the temperature change during a titration. A thermistor has advantages over a thermometer for measuring these small temperature changes because a thermistor can be very small and thereby have a very small heat capacity Furthermore, the resistance of a thermistor shows a large negative temperature coefficient and, therefore, temperature changes can be followed electrically by measuring the change of resistance with a Wheatstone bridge circuit. Automatic thermometric titrations may be performed by using a pen recorder for measuring the temperature from the unbalance of the Wheatstone bridge, together with continuous addition of titrant from a motor-driven syringe. *See* THERMOCHEMISTRY.

Even though thermometric titrations have been applied to the titration of weak and strong acids and to various precipitation and complex-formation reactions, the most useful applications for this technique will probably be in nonaqueous systems and in situations where it is desirable to perform analyses continuously. *See* CONCENTRATION SCALES; STOICHIOMETRY. [C.E.B.]

Titration, amperometric

A titration that involves measuring a current or changes in a current during the course of the titration (*see* TITRATION). From polarographic data on half-wave potentials for reductions or oxidations of various ions and molecules, it is possible to find a setting of applied electromotive force (emf) for an indicator electrode of dropping mercury or microplatinum versus a reference electrode that will be appropriate to give a diffusion current that is proportional to the concentration of one of the reactants in a titration. The observed current at the indicator electrode is then a direct measurement of the variation of the concentration of one or both of the reactants during the course of the titration, and from these measurements it is possible to determine the equivalence point.

The measured current may be corrected for the small residual current due to condenser effects and impurities in the solution being titrated. This, however, it not always necessary. A correction of the observed current for the volume change during a titration is invariably made and is done by merely multiplying the observed current by $(V + v)/V$ where V is the initial volume of the solution being titrated and v is the volume of titrant added. In amperometric titrations, the corrected current, i_{corr}, is plotted against v Some typical graphs for amperometric titrations are shown in the illustration. E.P. denotes equivalence point in each case. Graph *a* represents the case where the ion that is *being titrated or precipitated is reducible and the* ions in the reagent or titrant are not reducible. A



(a–d) End-point (E.P.) detection in amperometric titration.

ypical example would be the titration of a lead olution with sodium sulfate. Graph *b* illustrates a ration where the substance that is being titrated

mL. The titration of lead nitrate with potassium chromate is an example of this type. Graph *d* illustrates an oxidation reduction titration, namely, the titration of thallium(III) with iron(II). The descending branch above zero current is caused by ... substance thal-

beyond the equivalence point. The different slopes are caused by the different diffusion coefficients of the ions that are undergoing reaction.

As is apparent from the graphs, amperometric titrations use an extrapolation of the plotted data to obtain the equivalence point. This has the advantage that the current at the exact equivalence point does not have to be measured. Furthermore, because the solubility of a precipitated compound is always the maximum at the equivalence point (that is, there is no common ion present to depress the solubility as there is on either side of the equivalence point), the observed current is high in the vicinity of the end point (graph *c*). For this reason, any measurements made close to the equivalence point are usually not used in determining the slopes of the straight lines.

Amperometric titrations have the advantage that a small amount of a given substance may often be titrated accurately in the presence of a very large

electrode, and this amount is, in turn, proportional to the concentration of this substance in the bulk of solution. Because all ionic species contribute to electrical conductance, this situation is one that is unfavorable for conductimetric titrations. Also, there may not be suitable indicator electrodes for potentiometric titrations. For example, it is difficult to find good indicator electrodes for nickel, bismuth, or lead ions in potentiometry, whereas the dropping mercury electrode may be used as an indicator.

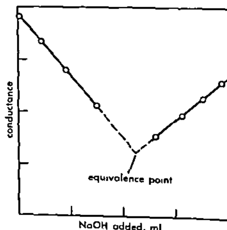
The apparatus for amperometric titrations can be very simple. The currents are measured with a galvanometer or a microammeter, and the desired emf may be obtained from any convenient battery connected across a potentiometer resistor. In some cases, the potential of the reference electrode will polarize the indicator electrode at the proper potential automatically when the two electrodes are connected through a resistance and a galvanometer in series. See ELECTROLYTIC CONDUCTANCE, POLAROGRAPHIC ANALYSIS.

[C.E.B.]

Titration, conductimetric

A titration in which the electrical resistance of a solution is measured during the course of the titration. Electrical resistance is merely the reciprocal of the conductance, and in conductimetric titrations, it is customary to refer to the conductance instead of the resistance of a solution. In many titrations, the concentrations and the mobilities of the ionic substances change during a titration. Because the electrical conductance of a solution is directly proportional to the number and the mobility of ions present, the conductance varies during these titrations. Therefore, by plotting the conductance as a function of the volume of titrant added, a conductimetric titration curve is obtained from which the equivalence point can be determined.

For example, if a solution of hydrochloric acid is being titrated with a standard solution of sodium hydroxide, the initial conductance of the analyte (hydrochloric acid solution) is a result of the presence of hydrogen or hydronium ions and chloride ions. When sodium hydroxide is added, hydrogen ions react with the hydroxyl ions and are removed from solution. At the same time, a concentration of sodium ions equivalent to the hydrogen ions removed are added to the solution. Because sodium ions have a lower mobility than hydrogen ions, the conductance of the solution decreases. Further addition of sodium hydroxide produces the same effect, until the equivalence point has been reached. Now, further addition of sodium hydroxide merely adds sodium and hydroxyl ions to the solution and the conductance increases. If the conductance measurements for this titration are plotted for various increments of titrant added, a curve similar to that shown in the illustration is obtained. In such titrations, it is not necessary to measure the conductance at the equivalence point. Generally, the conductance is measured for three or four definite increments of titrant before and after the equivalence point, and the equivalence point is determined from the intersection of the straight lines drawn through the experimental points.



End-point detection in conductimetric titration

Because the mobility of ions increases by ap-
roximately 2% for each centigrade degree rise in emperature, the temperature of the solution must emain virtually constant during a titration. This ondition is met by carrying out the titration in a essel suspended in a constant-temperature bath. lso, the addition of titrant dilutes the ions and hereby changes the concentration. If the titrant s 20–50 times as concentrated as the solution being trated, a correction for dilution effect may not be ecessary. However, one can be made by merely ultiplying the measured conductance by the ratio $V + v)/V$, where V is the initial volume of analyte nd v is the volume of titrant added.

Conductance measurements are usually made vith two platinum-foil electrodes which have been oated electrolytically with a thin deposit of plati-um black. These electrodes are maintained at a ixed distance apart and in a fixed location in the olution. The resistance between these electrodes is easured with a Wheatstone bridge circuit which s fed from an alternating current source, usually f 1000 cps

Conductimetric titrations are well adapted to the etermination of acids or bases, and to titrations in vhich precipitates or complexes are formed. The ethod is not useful for oxidation-reduction proc-sses unless there is a considerable consumption or roduction of hydrogen or hydroxyl ions during he reaction. If a highly ionized acid is titrated vith a highly ionized base, or the reverse, an accu-ate end point may be obtained even in very dilute olution.

Weak acids or bases may also be titrated accu-ately over a considerable concentration range. A uccession of end points is found if two acid or base unctions of very different strengths are present. For example, a mixture of hydrochloric and acetic cids or of acetic and boric acids may be titrated.

A single conductance measurement may serve for he determination of a particular substance, pro-ided suitable prior calibration has been made. For xample, a two-component mixture can frequently e analyzed by this technique In addition, the onductance of a solution of an organic substance s frequently directly proportional to the amount f ash that the organic substance leaves when urned. Thus, the ash of such compounds may be etermined from a single conductance measure-ent.

In addition to the more classical methods of mak-ng conductance measurements, it is possible to btain good results in titrations by using small onstant direct currents applied between two elec-rodes and then observing the IR drop with a vac-uum-tube voltmeter between two additional elec-rodes. Also, high-frequency titration techniques ave been introduced since World War II. With this echnique, two electrodes are mounted on the out-side of the beaker or vessel and an alternating cur-rent source in the megacycle range is used. Even though capacitance terms as well as ordinary con-ductance are measured with such circuits, it is pos-sible to utilize such measurements to follow the course of a titration. This technique is attractive because the electrodes are outside of the solution, and it is particularly applicable to measurements that are to be made on a continuously flowing liquid. Because the response of such equipment is extremely sensitive to changes in dielectric con-stant, a very practical application is the deter-mination of small amounts of water in organic substances. *See* ELECTROLYTIC CONDUCTANCE; TITRATION. [C.E.B.]

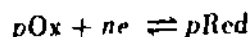
Titration, potentiometric

A titration the course of which is followed by meas-uring the potential of a suitable electrode. Because the potential of a single electrode cannot be meas-ured, a second electrode whose potential remains constant is used as a reference and the emf of the resulting galvanic cell is measured. This emf is usually measured by means of a potentiometer or a vacuum-tube voltmeter. The term potentiometric should be reserved for those measurements where no current is drawn from, or put into, the electrode system except momentarily while balancing the potentiometer to read the emf or except for the minute current required to operate the vacuum tube voltmeter.

Potentiometric indication of end points may be applied to any type of titration provided an elec-trode can be found that responds directly or in-directly to changes in concentration of one of the reactants The response of an electrode in volts at 25°C is determined from the Nernst equation and is $(0.059/n)$ log a_{ion} for reactions that involve com-binations of ions, as in neutralization, precipitation, and complex-formation A response of

$$\frac{0.059}{n} \log \frac{(a_{\text{ox}})^p}{(a_{\text{red}})^p}$$

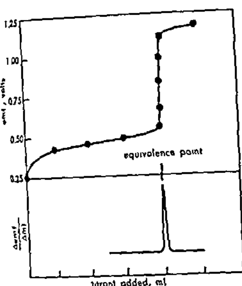
is obtained for oxidation-reduction reactions where n is the number of electrons exchanged per mole-cule or ion, a is the activity (or effective concentra-tion) of the reactant, and p represents the coeffi-cients in the partial reaction:



$$p\text{Ox} + ne \rightleftharpoons p\text{Red}$$

for a soluble oxidation-reduction couple

The potential of a piece of platinum wire or foil immersed in the solution may be used to follow the titration of many oxidation reduction titrations This potential is actually determined from the emf of the cell composed of the platinum electrode and a reference electrode, such as a saturated calomel electrode Because the potential of the reference electrode remains constant, any change in the plat-inum electrode is reflected in the emf measurement A potentiometric titration curve, which is a plot of the measured emf against the volume of titrant added, is shown in the illustration The equivalence point occurs at, or close to the volume correspond ing to the inflection point in the steep portion of the curve or where $\Delta(\text{emf})/\Delta(\text{ml})$ is a maximum or where $\Delta^2(\text{emf})/\Delta(\text{ml})^2$ changes sign and passes through zero



End-point detection in potentiometric titration

With a suitable indicator electrode and a reference electrode, each potential measurement indicates the activity of a chemical entity, provided liquid junction potentials are either eliminated or handled by appropriate corrections Such measure-

[illegible]

only to this single ion. Therefore, potentiometric measurements find wide application, not only in following titrations, but also in the direct measurement of the actual activity of certain chemical species in solution

Probably the widest application of potentiometric measurements is found in measuring hydrogen-ion concentration by means of pH meters Most readings with a pH meter involve changes in potential of a glass electrode which responds essentially only to hydrogen ions. These measurements can be used to follow the course of acid base titrations or merely to determine the hydrogen-ion concentration in a solution The response of a pH

[illegible]

unknown hydrogen ion concentration.

Potentiometric measurements are, in general, capable of high precision. Potentiometric titrations have the additional advantages that they are applicable in colored solutions they enable one to [illegible] end [illegible] be- [illegible] POTENTIAL, ELECTROMOTIVE FORCE (CELLS); HYDROGEN ION, OXIDATION REDUCTION; TITRATION [C F B]

Toad

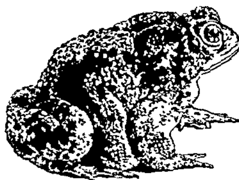
Any member of the amphibian family Bufonidae in the order Salientia (or Anura) This family is almost world wide in distribution, excepting only the colder climates and Australia Madagascar, New Guinea, and Polynesia. [illegible] genera, with the genus *Bufo* by far the largest, consisting of about 250 species, 16 of which occur in the United States. This genus has the same range as the family Bufonidae.

Toads are relatively short-legged, dry-skinned animals. The upper surface is covered with warty outgrowths supplied with glands which produce an [illegible] This

[illegible]

The mild irritant from these glands is apparently the origin of the superstition that handling toads will cause warts on one's hands. Large paired parotid glands, one located behind each eye, are prominent surface features.

Toads differ from frogs in remaining near water only during the breeding season. Large numbers of toads occur in desert regions They are further evolved toward freedom from the water by a shortened tadpole stage lasting only from 2 to 8 weeks, in contrast to the 1–3 years required for frogs.


The common toad, *Bufo americanus*, length male to 3½ in. (*From E L Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

Toads are nocturnal coming out of hiding at night to feed upon insects, worms, and other invertebrates They are considered highly beneficial animals because of their feeding habits

Toad eggs are produced in long, gelatinous strings The American toad, *Bufo americanus*, lays about 15,000 eggs each spring. Fertilization is external

Toads hibernate during the winter by burrowing into the soil. The American toad sometimes digs to a depth of 3 ft

The spadefoot toads, represented in the United States by six species and belonging to the family *Pelobatidae*, are nocturnal burrowing toads with thin, relatively smooth skins and vertical pupils See AMPHIBIA. [J D B.]

Tobacco

The plant genus *Nicotiana*, certain species in the genus, and dried leaves of these plants are all called tobacco. Most often tobacco means a leaf product containing 1–3% of the alkaloid nicotine which produces a narcotic effect when smoked, chewed or snuffed The plant *Nicotiana rustica* provides tobacco in parts of Europe, but the tobacco

of world commerce is *Nicotiana tabacum* (Fig. 1).

Tobacco is American in origin. Columbus found West Indians smoking it in a hollow forked stick. Historians do not know who first brought tobacco to Europe, but most of them credit Jean Nicot in 1561. Nicotiana and nicotine bear his name.

Characteristics. This solanaceous annual, found only in cultivation, probably began as an amphidiploid or fertile hybrid between two species of *Nicotiana* native to Bolivia (*see* GENETICS). It has 24 chromosome pairs, an erect, thick stem 2–9 ft high, and alternate, sessile, oval, or lanceolate leaves. The flowers are produced in a panicle (*see* INFLORESCENCE). They have a tubular corolla 3.5–5.5 cm long widening midway to a throat and ending in a five-pointed flare (*see* TUBIFLORALES). There are five anthers which shed sticky pollen just as the flower opens. Most flowers are self-pollinated. The fruit is a two-section capsule with minute seeds (about 15,000 in a gram).

Cultivation, harvesting, and curing. Tobacco generally does best on light, well drained, and carefully fertilized soils that are clean-cultivated, that is, free of undergrowth, and receive moisture weekly. Harvest proceeds by cutting the whole stalk (stalk-cut) or picking leaves successively as they ripen (primed). Primed leaves are supported on strings, wires, or sticks. Drying is done with natural or artificial heat. This is called curing. Drying time (½–6 weeks) and temperature (70–170°F) influence the amount and kind of changes that occur in proteins, carbohydrates, organic acids, alkaloids, and enzymes in the leaf. Before use in cigars, cigarettes, pipes, chewing tobacco, or snuff, cured leaves are fermented by storing them 6 weeks to 2 years at about 15% moisture and 80–110°F. The methods used for harvesting, curing, and fermenting depend on the type of tobacco, intended use, and local custom.



Fig. 1 *Nicotiana tabacum*. Connecticut cigar wrapper type (*Connecticut Agriculture Experiment Station*)

Economic importance. Tobacco is economically important in 66 countries and is grown to some extent in all but a few countries. In the United States, the average annual farm value of tobacco from 1945–1954 was $1,025,789,508. Data on production and farm value for United States tobacco types in 1956 are shown in the table. *See* AGRICULTURAL SCIENCE (PLANT) [G.S.T.]

Major United States tobacco types and 1956 production

Type	Where grown	Type of cure	Use	1956 Production and farm value		
				1000 lb	Acres	$1000
Flue-cured (bright)	Virginias, Carolinas, Georgia-Florida border	Picked leaves heat-cured to bright yellow	Cigarettes	1,422,538	875,200	732,607
Burley	Tennessee, Kentucky	Whole plant air-dried in ventilated sheds	Pipe and cigarette blends	506,395	309,800	321,560
Cigar filler	Pennsylvania, parts of Ohio and New York	Same as above	Central bulk of cigars	57,600	34,000	13,708
Cigar binder	Connecticut River Valley, Wisconsin	Same as above	Binding central bulk into cigar shape	33,970	19,100	13,014
Cigar wrapper	Under cheesecloth cover in Connecticut River Valley and Georgia-Florida border	Picked leaves heat- and air-dried to a golden brown	Outer leaf or wrapper of cigar	17,162	13,300	33,809

acco diseases. This term is used to designate ety of abnormalities of the tobacco plant that ely affect its value. Diseases result from both ic and inorganic causes.

eather. Excessive cold causes chlorosis, or loss lorophyll, in seedling leaves (*see* CHLORO- ; PHOTOSYNTHESIS). High temperature and ant light cause breakdown of chlorophyll in r leaves. Beating rain bruises the upturned rsides of leaves. Hail tears the leaves and ses the stems.

utritional disturbances. A soil acidity of about 4 releases manganese in concentrations toxic leaves. Mineral deficiencies cause many pro- nced abnormalities. For example, a deficiency

po

eat

tro

the lower leaves, ...

d late maturity; of boron, by necrosis of the rowing point; of calcium, by irregularly shaped eaves. Frenching, or the crinkling or rolling of eaves, is associated with a bacterial toxin and an mbalance of amino acids (Fig. 2). *See* PLANT, MINERALS ESSENTIAL TO.



. 2. Frenching, a physiological disease of tobacco

Bacteria Slime disease, *Pseudomonas* (*Bacillus*)

hrough stomata into water-congested tissue where a colony develops Black leg and hollow stalk, *Erwinia* (*Bacillus*) *aroideae*, are soft stem rot of seedling and pith rot of mature plants, respectively. Bacterial black stalk is caused by an unidentified organism. *See* BACTERIA; ENTEROBACTERIACEAE; PSEUDOMONADACEAE

Fungi Black root rot, *Thielaviopsis basicola*, and black shank, *Phytophthora parasitica* var *nicotianae*, are destructive root diseases of cool and warm areas, respectively (Fig 4) Blue mold, *Peronospora tabacina*, is a leaf disease destructive to tobacco seedlings in the southeastern United

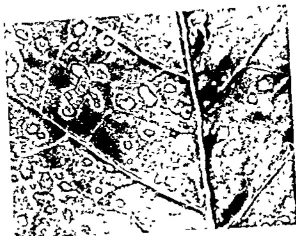


Fig. 3. Wildfire, a bacterial disease of tobacco.



Fig. 4 Black root rot of tobacco (a) A susceptible plant (b) A resistant plant. Both plants were grown in soil infested with the black root rot fungus

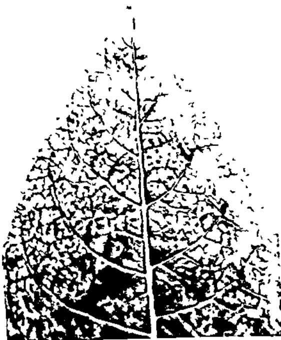


Fig 5 Mosaic, caused by the tobacco mosaic virus.

States and Australia. Powdery mildew, *Erisyphe cichoracearum*, is destructive in most tobacco-growing areas except the United States. Fusarium yellows and other minor leaf and root diseases are caused by fungi. *See* FUNGI.

Viruses. Viruses attacking tobacco are mosaic (Fig. 5), etch, ring-spot, cucumber mosaic, vein-banding or potato Y, leaf curl, spotted wilt, curly dwarf, clubroot, streak, rattle, yellow dwarf, big bud, and others. Mosaic is spread by man and tools; the others are probably spread by insects. All have wide host ranges and are usually found in native vegetation. *See* PLANT VIRUS.

Nematodes. Several genera of nematodes live either internally in root knots or lesions or feed externally on the tender roots. Root damage is extensive in sandy soils. *See* NEMATODA.

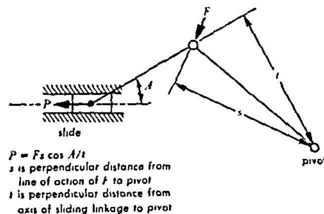
Parasitic plants. Broomrape (*Orobanche* sp.) and witch weed (*Striga* sp.) are parasitic on tobacco roots, whereas dodder (*Cuscuta* sp.) attacks seedlings and sometimes field plants.

Humidity during curing. Air-cured tobaccos are damaged by exposure to an average relative humidity much above 65%. *See* HUMIDITY

Disease control by breeding. Resistance to mosaic, black shank, powdery mildew, wildfire, blue mold, and black root rot have been transferred from wild species to cultivated tobacco. Resistance to slime disease has been found in *N. tabacum* and introduced into flue-cured tobacco. *See* PLANT DISEASE; PLANT DISEASE CONTROL. [W.D V.]

Toggle

Two joined linkages with the far end of one pivoted about a fixed point and the far end of the other sliding along a line (see illustration). A toggle joint is used as a snap-action mechanism in electric switches, the pivoted linkage being extended beyond the pivot to serve as an actuating handle. In a toggle joint, a small force at the junction of the two linkages produces a much larger force at the slider; therefore, the mechanism is used in presses and crushers *See* LINKAGE, MECHANICAL. [F.H R.]

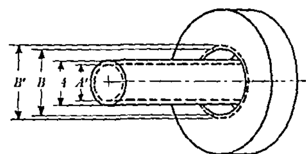


$P = Fs \cos A/t$
s is perpendicular distance from line of action of F to pivot
t is perpendicular distance from axis of sliding linkage to pivot

Basic toggle mechanism and its characteristic equation.

Tolerance

Amount of variation permitted or "tolerated" in the size of a machine part. Manufacturing variables make it impossible to produce a part of exact dimensions; hence the designer must be satisfied with manufactured parts that are between a maximum size and a minimum size. Tolerance is the difference between maximum and minimum limits of a basic dimension. For instance in a shaft and hole fit, when the hole is a minimum size and the shaft is a maximum, the clearance will be the smallest, and when the hole is the maximum size and the shaft the minimum, the clearance will be the largest (see drawing).



—— size for basic dimension
- - - - size with tolerance
A = basic diameter of shaft
B = basic diameter of hole
$B - A$ = allowance
$B' - A'$ = maximum clearance
$A - A'$ = tolerance on shaft
$B - B'$ = tolerance on hole

Shaft and hole dimensions

If the initial dimension placed on the drawing represents the size of the part that would be used if it could be made exactly to size then a consideration of the operating conditions of the pair of mating surfaces shows that a variation in one direction from the ideal would be more dangerous than a variation in the opposite direction The dimensional tolerance should be in the less dangerous direction This method of stating tolerance is called unilateral tolerance and has largely displaced bilateral tolerance in which variations are given from a basic line in plus and minus values

As an example, for a 1½-in. shaft and hole for a free fit the standard allowance is 0.002 in., and the tolerance for hole and shaft is 0.001 in

Maximum shaft diameter = nominal size − allowance
= 1.500 − 0.002 = 1.498 in

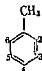
In the unilateral method for stating tolerance the shaft diameter is $1.498^{+0.000}_{-0.001}$ or between 1.498 and 1.497 in The diameter of the hole is $1.500^{+0.001}_{-0.000}$ or between 1.500 and 1.501 in The maximum clearance

is 1 501 minus 1 497, or 0 004 in., and minimum clearance is 0 002 in. [P.H B]

Bibliography: E. Buckingham, *Production Engineering*, 1942; E. Oberg and F. D. Jones, *Machinery's Handbook*, 15th ed., 1954.

Toluene

A colorless, aromatic hydrocarbon, also called methylbenzene, which boils at 110 6°C and freezes



at −95 0°C. Most of the toluene sold in the United States is produced by the action of catalysts on petroleum hydrocarbons.

The nucleus of toluene, like that of benzene, undergoes substitution reactions. Substitution occurs almost exclusively in the ortho (2) and para (4) positions. Nitration under vigorous conditions yields 2,4 6-trinitrotoluene (TNT), an explosive. The hydrogen atoms of the —CH_3 group may be replaced by halogen at high temperatures in the presence of ultraviolet light.

Toluene is an important ingredient in high-octane gasoline, a solvent for gums and lacquers, and an intermediate in the manufacture of TNT, benzaldehyde, and benzoic acid.

[illegible]

HYDROCARBON; BENZENE [C.K B.]

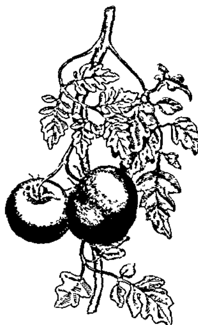
Toluidine

One of three organic isomeric chemical compounds, $CH_3C_6H_4NH_2$, that are homologs of aniline, with chemical properties very similar to those of aniline. The ortho derivative boils at 202°C, the meta at 45°C, and the para melts at 45°C Nitration of toluene yields all three nitrotoluenes (meta in small amount), and subsequent reduction of the separated isomers with iron and water in the presence of acid gives the corresponding amines (toluidines). The toluidines are weak bases of about the same strength as aniline. All are used to make azo dyes and rubber antioxidants. *See* AMINE; ANILINE, NITROBENZENE. [L.B C]

Tomato

Species of the genus *Lycopersicon*, especially *L esculentum*, a widely cultivated vegetable grown for its edible fruit. The genus probably originated in the Peru Ecuador area of South America. Among the many names commonly used for this plant are pomodoro and loveapple. The tomato was domesticated in Mexico before the discovery of America, and it was brought to Europe early in the period of the Spanish conquest However, it was not until the eighteenth century that the tomato became generally used for food. The United States tomato industry became firmly established during the latter half of the nineteenth century.

Classification and structure. The genus *Lycopersicon* belongs to the plant order Tubiflorales, as do the potato, eggplant, and pepper (*see* TUBIFLORALES). The genus is divided into two subgenera—the red-fruited *Eulycopersicon*, including *L. esculentum* and *L. pimpinellifolium*, and the green-fruited *Eriopersicon*, including *L. cheesemanii*, *L. peruvianum*, *L. hirsutum*, and *L. glandulosum*. Another species, *L pissisi*, is sometimes included. The haploid chromosome number of the genus is 12 (*see* GENETICS). These species are closely related as to morphology, environmental reactions, and cytology (*see* CYTOLOGY; PLANT ANATOMY). Although crosses of some species are difficult to make, nearly all interspecies crosses have been achieved, and these have been sufficiently fertile to permit experimental studies. In temperate regions, the tomato is grown as an annual (*see* ANNUAL PLANTS).



Tomato, *Lycopersicon esculentum* (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

The growth habit varies from a spreading vine to a semierect or erect plant. Adventitious roots develop readily from the stem. *See* ROOT (BOTANY); STEM (BOTANY). The leaves are odd-pinnate with seven to nine short-stemmed leaflets The arrangement is usually alternate with a 2/5 phyllotaxy. *See* LEAF (BOTANY). The inflorescence is made up of a raceme of 4–12 flowers (*see* INFLORESCENCE). The yellow flowers are perfect and hypogynous. The 5 stamens with short filaments are united laterally to form a hollow cone around the pistil. The pistil, consisting of two to several carpels, extends through the anther cone. *See* FLOWER (BOTANY). The red to yellow fruit is a two- to many-loculed berry with fleshy placentas containing many small

oval seeds covered with short hairs. *See* Fruit (botany); Seed (botany). The tomato is mostly self-pollinated although various ecological conditions may result in a considerable amount of cross pollination. *See* Reproduction, plant.

Distribution and economic importance. The tomato is now cultivated in all parts of the world. The leading countries in terms of acreage are the United States, Italy, Mexico, Egypt, and Brazil. In 1955, approximately 1,600,000 acres were under cultivation in the world of which 500,000 were in the United States. From 1950–1955, the United States tomato crop had an average yearly farm value of $218,000,000, of which $128,000,000 resulted from fresh-marketing and $90,000,000 from processing. The tomato industry can be divided into greenhouse (forcing), fresh-market (truck), and canning (processing). The leading states in production for fresh-market are Texas, Florida, and California. The leading tomato-processing states are California, Indiana, and New Jersey. Tomatoes represent a major portion of the processing industry. *See* Food manufacturing; Vegetable growing.

The tomato is grown in many different soils and under a variety of climatic conditions. It is classified as a warm-season plant because it cannot withstand frost. The usual method of planting is with transplants (seedlings), although tomatoes for canning may be grown directly from seed.

Tomato pests. The tomato is subject to many pests including fungi, bacteria, viruses, nematodes, and insects (*see* Bacteria, Fungi; Insecta; Nematoda; Plant virus). Many of the ravages of these diseases can be overcome by plant protection and disease control (*see* Plant disease control). The incorporation of natural resistance by suitable breeding methods has resulted in many resistant varieties.

Culinary and biological uses. The popularity of the tomato is largely due to the great variety of its culinary uses. raw or cooked It is a good source of ascorbic acid (28 mg per 100 g of fruit). The fruit has a low pH (4.25–4 30) and is relatively easy to process. Tomatoes are canned whole and as juice, puree, sauce, catsup, and paste.

The tomato has become a favorite test plant for a wide range of biological investigations because of its ease of propagation by seed and by cuttings and its adaptability to a wide range of environmental conditions. It has been used extensively for plant physiological studies including the control of flowering and abscission and plant nutrition (*see* Plant physiology). The genetics and cytology of the tomato have undergone much study and, next to corn, it is one of the most extensively investigated plants. [N.K.E.]

Bibliography: *See* Agricultural science (plant).

Ton of refrigeration

A rate of cooling that is equivalent to the removal of 200 Btu of heat per min, 12,000 Btu/hour, or 288,000 Btu/day. This unit of measure stems from the original use of ice for refrigeration. One po of ice, in melting at 32°F, absorbs as latent approximately 144 Btu/lb, and one ton (2000 of ice in melting in 24 hours absorbs 288 Btu/day. In Europe, where the metric systen used, the equivalent cooling unit is the frig which is a kilogram calorie, or 3.96 Btu. Thus 3 frigories/hour is approximately one ton of ref eration. A standard ton of refrigeration is one veloped at standard rating conditions of 5°F e orator and 86°F condenser temperatures, with liquid subcooling and 9°F suction superheat. Refrigeration. [H.M.

Tonalite

A phaneritic (visibly crystalline) plutonic r composed chiefly of plagioclase (oligoclase or desine) and quartz with subordinate dark-colo (mafic) minerals (biotite, amphibole, or pyr ene). The term tonalite is roughly equivalent quartz diorite Minor amounts of alkali felds may be present, but if this mineral exceeds 5% the total feldspar, the rock is a granodiorite. As quartz content decreases, quartz diorite passes i diorite. Tonalite, or quartz diorite, is roughly int mediate between granodiorite and diorite. *See* orite. [C.A.C

Tone (music and acoustics)

A sound oscillation capable of exciting an audito sensation having pitch; also, the sensation itse that is, the word tone is used for both cause a effect. There is not necessarily a complete cor spondence between the two; which of the t meanings is intended must be made clear by ad tional modifiers, context, or units of measureme For example, if a tone is described as having pitc it is to be understood that the sound sensation meant, whereas a tone that has frequency must be physical oscillation. *See* Oscillation; Pitch

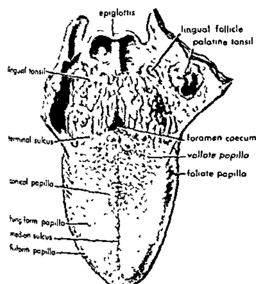
The word note is often used as a synonym f tone, either as the sensation or the oscillati causing the sensation, even though a note is pr marily a symbol to indicate the pitch and durati of a tone sensation. Thus, note serves when no di tinction is desired among the symbol, the sensatio and the physical stimulus

Tone is also an interval in music such as betwee the first and second notes of an ordinary maj scale; the ratio of the two frequencies forming th interval is approximately the sixth root of two. Th interval is called a step, or whole step, an interv half the size is the semitone or half-step, the rat in this case being approximately the twelfth roo of two. *See* Musical acoustics [R.W.Y.

Tongue

An organ located in the floor of the mouth. I higher vertebrates the tongue is mobile because o contained muscle; fishes have a so-called tongu that is merely an elevation on the floor of the mouth lacking muscle. Amphibians vary from the absence of a tongue or one similar to a fishlike condition t

dilians lack a protrusible tongue, but snakes and lizards have a highly protractile tongue which is usually forked. The tongue of birds, usually covered with a horny substance and sometimes bearing barbs, is practically lacking in intrinsic muscles. Woodpeckers are able to extend the tongue remarkably, but this is because the hyoid bone moves forward and pushes the tongue before it. The tongue reaches its highest development in mammals where it is specialized in form and function. Of the mammals, only whales lack a movable tongue with well-developed intrinsic muscles. The greatest mobility exists in the anteaters. The basal region of the mammalian tongue is probably like that of reptiles and birds; the fleshy superstructure is a new feature.



The tongue, dorsal view. (From M. W. Woerdeman, Atlas of Human Anatomy, vol. 2, Blakiston Division, Graw-Hill, 1950)

Mammalian tongue. The mammalian tongue is ided into two parts by a V-shaped groove, the minal sulcus. At the apex of this V is a small nd pit, the foramen caecum. It marks the point ere the thyroid gland arose as a bud and later parated. The larger part, or body, of the tongue longs to the floor of the mouth, whereas the remainder, or *root*, forms the front wall of the oral larynx. The body of the tongue is separated from e teeth and gums by a deep groove. A midline ld, the frenulum, is *near the tip* on the under surce. The upper surface of the body, called the orsum, has a velvety appearance because it is ickly studded with rows of *tiny, tapering filiform* apillae. Distributed among these are occasional rger, rounded fungiform papillae and some large onical papillae. Immediately in front of the groove separating the body of the tongue from the root is a series of still larger vallate papillae arranged in a V-shaped row. The apex of the V points down the throat. Posteriorly along each side of the body of the tongue and near the root, is a series of parallel folds constituting the foliate papillae. The surface of the root of the tongue, which belongs to the pharynx, is devoid of papillae but bears warty nodules containing lymphoid tissue.

Anatomically the mammalian tongue is a mass of skeletal muscle largely contained within a covering sheet of mucous membrane. The surface epithelium is of the stratified squamous type. It is cornified in some mammals but not in man. Filiform papillae contain a core of connective tissue subdivided so that the whole papilla resembles a cat-o'-nine-tails. They are best developed in the cat family where they convert the tongue into a rasping organ. Fungiform papillae are more prominent elevations, with a knobbed top somewhat like a button mushroom. Some bear a few taste buds. Conical papillae are considered to be a modified fungiform type. Vallate papillae number 7–11 in man. They have a flat top, 1–3 mm in diameter, and do not extend much above the lingual surface. Each is encircled by a relatively deep trench. The side surface of a human vallate papilla contains some 200 taste buds, whereas the opposite wall, across the trench, bears about one-fourth as many. These numbers are quite variable, however, and many buds disappear with age. Foliate papillae are well represented in rodents. Human infants have 4–8 of these prominent folds on each side of the tongue which bear numerous taste buds; this series of folds becomes regressive in adult man.

Taste buds. Taste buds are quite similar in all vertebrates. Each is an ovoid specialization paler *than the surrounding stratified squamous epithelium*, they occupy practically the full thickness of that covering layer. The shape is somewhat like a barrel, *but the top usually narrows. Two cell types* are recognizable within the taste bud. The taste cells are slender, spindle-shaped elements whose free end terminates in a short, stiff taste hair. These hairs extend into the taste canal, hollowed out from the more superficial layers of the epithelium, and continue to the external taste pore. The supporting cells of the taste bud are mostly at the periphery, something like thick barrel staves. Nerve fibers for the mediation of taste end about both kinds of cells.

The mucous membrane that covers the root of the tongue differs from that of the body of the tongue and it is characterized by much lymphoid tissue. The bumpy appearance of the surface is caused by many epithelial pits (35–100 in man) about which lymphocytes mass to produce a mound; some of these are several millimeters in diameter. The aggregate of these units comprises what is known as the lingual tonsil.

Lingual glands. Lingual glands lie deep in the mucous membrane of the tongue and encroach on the muscle beneath. Mixed muco-serous glands occur *under the apex of the tongue*. Purely serous

glands of Von Ebner are restricted to the region of the vallate and foliate papillae and many discharge on the surface of the tongue. Other glandular ducts open into the trench of each vallate papilla. The root of the tongue contains pure mucous glands, many of which open on the surface but some of which have ducts discharging into the pits of the lingual tonsil.

Muscle. The mass of muscle that comprises almost all of the tongue is halved incompletely by a median lingual septum of fibrous tissue. Its course is indicated on the surface of the tongue by a longitudinal median sulcus. The voluntary muscle fibers belong to two groups: intrinsic fibers lie wholly within the limits of the tongue, extrinsic fibers enter from without and also serve to anchor the tongue. The muscle fibers are arranged in definite bundles that interlace at right angles. Some bundles are longitudinal in direction; others pass vertically and still others run horizontally.

Function. The mammalian tongue performs a variety of functions. In ruminants it is prehensile and thus is important in browsing. Some mammals use the tongue to lap up liquids. In general, it is of use in directing food to the teeth and pharynx, aiding chewing and swallowing. The dog, lacking sweat glands, cools itself by panting, which draws air over the tongue. The tongue is the most important agent in articulate speech. The sense of taste resides in the taste buds, whose cells respond to substances in solution.

Development. The body of the mammalian tongue develops from a pair of swellings on the first branchial arches that also become the lower jaw. The line of union of these primordial components is marked by the median sulcus. A median mass, the tuberculum impar, between the first and second arches becomes compressed and contributes little, except the septum, to the body of the tongue. The root of the tongue arises chiefly from the union of the ventral ends of the second pair of branchial arches, but the third and fourth arches apparently contribute as well. The junction of body and root is indicated by the permanent terminal sulcus. The muscular component of the tongue is related historically to tissue that lay at the base of the head and migrated into the tongue, carrying with it the hypoglossal, or twelfth, cranial nerves. [L B A]

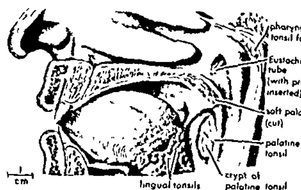


Fig. 1. Dissection showing nasal cavity above, oral cavity below, pharynx right (From J. Sobotta, J. McMurrich, ed., Atlas of Human Anatomy, vol. 2, 3 rev. English ed., Hafner, 1933)

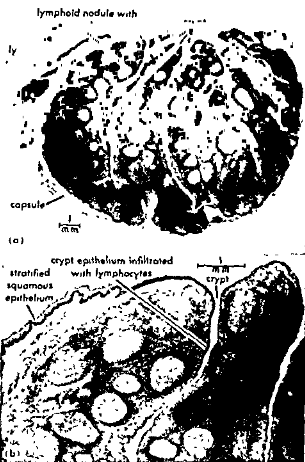


Fig. 2 (a) Vertical section of child's palatine tonsil, photomicrograph (b) Higher magnification of a

Tonsil

Localized aggregation of diffuse and nodular lymphoid tissue found in the region where the nasal and oral cavities open into the pharynx. The lymphoid tissue consists of small, closely-packed round cells called lymphocytes supported in a specialized connective tissue framework called reticular tissue. When lymphocyte production is active, rounded, more densely-packed clusters or nodules of these cells appear in the diffuse lymphoid tissue. The most active nodules possess lighter staining centers composed of somewhat larger, less densely-packed lymphocytes showing evidence of cell division. Such areas are called germinal centers. In the tonsillar regions, the lymphoid tissue lies just beneath the lining epithelium. The tonsils are important sources of blood lymphocytes. They frequently become inflamed and enlarged, thus necessitating their surgical removal.

Palatine tonsil. The two palatine (faucial) tonsils are almond-shaped bodies measuring 1×0.5 in and are embedded between folds of tissue connecting the pharynx and posterior part of the tongue with the soft palate (Fig. 1). These are the structures commonly known as the tonsils. The

openings of 10 to 20 pits (crypts) which extend deep into the organ may be seen on the surface. The stratified squamous epithelium of this region covers the surface of the tonsil and lines the crypts (Fig. 2). A fibrous capsule separates the tonsil from the underlying muscle. Extensions from the capsule form supporting septa within the tonsil. Lymphoid tissue occupies all interstices between the capsule, septa, and epithelium. Crypts frequently become filled with detached epithelial cells, living and dead lymphocytes, and exuding fluids. Such sequestered masses form an excellent culture medium for the growth of certain bacteria and fungi. The protective quality of the crypt epithelium may be weakened by the passage of large numbers of lymphocytes through it.

Lingual tonsil. The lingual tonsil occupies the posterior part of the tongue surface. It is really a collection of 35–100 separate tonsillar units, each having a single crypt surrounded by lymphoid tissue. Each tonsil forms a smooth swelling 2–4 mm in diameter. Since gland ducts open into these crypts, the contents are flushed out, and lingual tonsils rarely cause trouble. The epithelium is again stratified squamous, and a thin capsule is present around each unit.

***Pharyngeal* tonsil.** The pharyngeal tonsil (called adenoids when enlarged) occupies the roof of the nasal part of the pharynx and is covered with pseudostratified ciliated columnar epithelium. The organ consists of a series of radiating folds

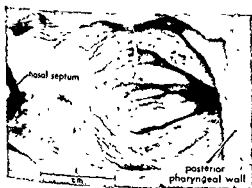


Fig. 3. Photograph of the roof of the nasal pharynx of a new-born baby showing the pharyngeal tonsillar folds.

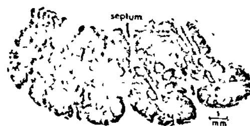


Fig. 4. Section across the pharyngeal tonsil folds of a child, photomicrograph.

leading forward from the region where the roof of the nasal pharynx joins the posterior pharyngeal wall (Fig. 3). When cut at right angles to the folds, the intervening spaces resemble crypts. Septa are found in the folds, but a distinct capsule is lacking. The lymphoid tissue and invasion of the epithelium by lymphocytes are similar to those of the other tonsils (Fig. 4). This tonsil may enlarge to such an extent that the nasal passageway may be blocked, forcing mouth breathing.

Function. *All tonsils produce lymphocytes* which are added to the circulating blood via the plexus of lymph capillaries which surrounds the lymphoid tissue. The flow of lymph is always away from the

a complete ring around the upper reaches of the digestive and respiratory systems. Because of this strategic location, a protective function has been suggested. It is thought by some that protection, in response to entering bacteria, may be afforded through the production of antibodies. Recent evidence has shown that plasma cells (many of which *are present in the tonsils), lymphocytes or both are* implicated in this process.

Development. Lymphocytic infiltration for the palatine tonsil begins at the site of the disappearing second pharyngeal pouch during the third fetal month. The pharyngeal and lingual tonsils appear during the fourth and fifth fetal months respectively. *Tonsils reach their maximum size during* childhood and subsequently regress. *See* HEMATOPOIESIS, LYMPHATIC SYSTEM. [T.S.]

Tonsillitis

An inflammation of the tonsils. The tonsils, being lymph tissue, readily become infected when they function in combating any intruding organisms to which they are exposed because of their location at *the entrance to the respiratory and intestinal* systems. Their deep epithelial pouches favor the growth of bacteria and the result is an inflammation, with the swollen tonsils causing a narrowing of the throat with painful swallowing, or angina. In order to fight infection, white blood cells penetrate into the pouches, where pus plugs form and can be seen as yellow dots at *the surface of the* organ. In some cases the epithelium which lines the *surface is destroyed and superficial* ulcers occur. In diphtheria these are extensive. The whole region is covered with a grayish membrane of leucocytes, fibrin, destroyed tissue, and bacteria.

The deep pouches favor the lodgment of bacteria and the development of chronic tonsillitis. These tonsils are a chronic focus of infection, where microorganisms and their toxins can *continuously* spread into the body. They represent one cause for *rheumatism. In children hypertrophy* of the lymphoid tissue leads to extreme enlargement of the tonsils, which can severely occlude the entrance to the throat. *See* PHAGOCYTOSIS; TONSIL. [E.WE.]

Tooling

Auxiliary devices used in manufacturing operations to supplement basic standard and special machine tools. Tooling is used, for example, on drill presses, milling machines, planers, and shapers to facilitate the actions of workers and to adapt the machine tools to the production of specific parts.

Tooling nomenclature. The over-all category of tooling encompasses a broad spectrum of manufacturing accessories; tooling is further subdivided into classifications such as jigs, fixtures, dies, gages, devices, gadgets, and other less universal designations.

There are no universally accepted definitions for differentiating between these subdivisions of tooling. Nomenclature for similar tools varies from one industry to another; it also varies between similar industries using similar devices. In shop practice, tooling subdivision nomenclature loosely follows the generalities of the accompanying table.

Types of tooling

Machine type	Prevailing tool nomenclature
Drill press	Drill jig
Boring machine (horizontal or vertical)	Fixture
Milling	Fixture
Welding	Jig
Assembly operations	Jig
Inspection operations	Gage
Presses (stamping, forming, and drawing)	Dies

Common hand tools and devices such as hammers, wrenches, pliers, screwdrivers, measuring scales, calipers and standard micrometers are normally excluded from the tooling category.

Standard perishable tools (consumed as a result of production operation) such as drills, reamers, milling cutters, and carbide tool bits are similarly excluded from the tooling category.

Special perishable tools, for example, a special-size drill, step drills, form cutters, and special tool holders, are normally regarded as tooling for a specific operation and are usually referred to as special cutting tools.

Auxiliary tooling. The option of physical separation of a machine tool and its tooling is not necessarily directly related to whether the machine tool is standard, semispecial, or special. There are valid reasons why all machine tools to a varying degree utilize removable tooling.

It is usually less costly and more practical to make the machine tool and the tooling (jigs, fixtures, dies) as entities rather than to have the same tooling features built directly into the body of the machine tool. For example, it may be most advantageous to make the body of the machine tool from cast iron; it may be more economical or mechanically necessary to make the tooling from a weldment, or the tooling device may best be a bolted assembly. The machining of complex details into a bulky machine body may be extremely awkward as



Fig. 1 Trunion-type drill jig in a radial drill press. In this operation the work is manually indexed so as to properly position the work to the drill. (American Tool Works)

compared to the same detail machining (tool making) of a relatively small tooling component.

By removing one set of tooling and replacing it with another (usually referred to as set-up) the usefulness of a machine tool (standard, semistandard, or special) can be extended so as to perform useful work on a variety of parts. Thus removable tooling is often an essential prerequisite for justifying procurement of, and for improving use of, a machine tool.

The use of removable tooling is a form of insurance against machine obsolescence due to product change. Existing tooling can be modified or new tooling created to cover unforeseen new product requirements.

Tooling can be modified or the effects of damage or wear repaired while the basic machine tool is doing other useful work without loss of productive capacity other than the change-over interval.

Functional uses. In operation, most tooling simultaneously performs some combination of a variety of purposes; the following represent only a few of the many purposes for which tooling is used: locating, clamping, positioning, cutter guiding, and others.

Locating. The locating function is accomplished by designing and constructing the tooling device so as to bring together the proper contact points or contact surfaces between the work piece and the tooling, and defining a direction of clamping force so that the work will infallibly assume the desired relationship to the tooling, and in turn thus have the desired relationship to the body of the machine tool and to the cutter.

There are usually three basic locating problems: concentric locating, plane locating, and radial locating.

Specific designs may require solution of only

by considering a simple analogy, that of locating a record on a phonograph turntable.

The hole in the record is placed down over the center pin of the turntable. This is known as concentric locating The center line of the record then coincides with the center line of the turntable. In this instance the record is the female member and the turntable pin is the male member of the concentric engagement. An actual tooling device may be the counterpart of either the record or the turntable, depending upon the part to be located in the tool

The record is now properly concentrically located, but it may still wobble in a horizontal plane unless it comes home against the face of the turn table. This is known as plane locating. In this instance the clamp is gravity which keeps the record in both plane and concentric relationship to the table.

A record does not require radial locating; however, suppose that the record must assume an exact radial relationship to the turntable, like the relationship of the winning number on a wheel of chance to the pointer finger. The record must then be so rotated

record in this relative position. This is known as radial locating

Empirical procedures in the designing of a tooling device are used because of the conflict between the many factors involved. At the specific stage of manufacture that requires a tooling device,



Fig. 2 Drill jig in which a hardened bushing guides the drill bit After first of two in-line holes is machined, jig is indexed 180 degrees and slip ring bushing moved to opposite liner (American Tool Works)

the work may already have different degrees of dimensional variations; for example, it may be rough, semifinished, or finished.

Work may be rough (such as castings, forgings, or weldments) and thus have considerable inherent dimensional variation.

Work may be semifinished; that is, it may be rough on some surfaces but accurately machined on others. There may be considerable variation between the rough and finished surfaces due to chucking variances in the previous machining operations.

Work may be finished all over, with negligible dimensional variation, or it may have appreciable variations between the various finished surfaces, particularly if they were machined in different independent operations.

Locating mechanisms that are best suited to rough locating are not usually adaptable to semifinished or to finished locating requirements and vice versa It is not a problem of degree but sometimes requires a different locating means for these different conditions.

Thus the combination of requirement for concentric locating, plane locating, and radial locating as well as the variances of rough, semifinished, and finished surfaces, and cost and time considerations, make locating a procedure requiring intimate first-hand knowledge of manufacturing operations. Sometimes it is essential to have a knowledge of the end use of the products, for this often inflicts restrictions or requirements that are not or perhaps cannot be practically expressed in the form of dimensional tolerances on the product drawing.

Clamping After the work is located it is clamped, this is a subsequent and cumulative problem. Clamps must act in the required direction with a proper degree of holding force.

In some respects a tooling device acts as a vise or as an anvil The tooling device must hold the work securely so as to resist the machining forces which may be considerable, as reflected in the amount of torque and pressure that is required to perform the necessary work such as cutting or deforming.

In holding the work securely, however, the clamp must not in itself deform the work to an objectionable degree The locating and clamping schemes should thus be contrived so that the operational forces do not act directly against the clamp (which might then require a powerful clamp) but rather so that the cutting forces are absorbed between the engagements of the work piece and the tooling, then requiring only a nominal amount of clamping pressure to properly hold together the part and the tooling

Positioning. The positioning functions of a tooling device may have a variety of interpretations and a variety of means to accomplish these ends.

, tilt, slide,
tionship to

The tooling may be required to position different components relative to each other for the pur-

pose of assembly or welding. In welding, the tooling may also be required to manipulate the work so as to achieve down-hand welding on different sides of the work.

The positioning function of the tooling device may be essentially a work-handling device where the work is heavy or otherwise impractical to handle by hand. The work itself may be in the hand-handling weight range, but the combination of the work and the necessary tooling may require mechanical assistance in the form of leverage, balance, or power actuation.

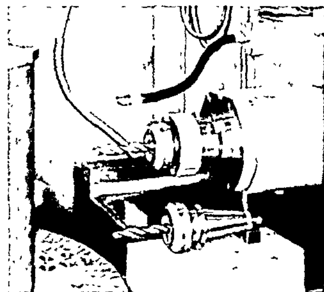


Fig. 3 Quick-change rigid tool holder for plunge drilling without guide bushing. For maximum rigidity, drill bit is stub held to project only minimum required distance. (*Reliance Electric and Engineering Co.*)

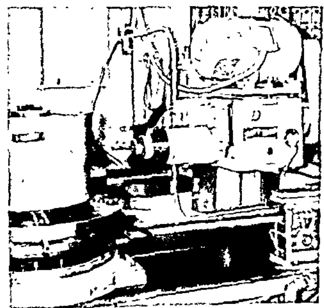


Fig. 4. Photograph of a plunge drilling operation. In this instance the work-holding device is mounted on a master index table which permits drilling of various combinations of radial holes into the tubular work piece. The horizontal drill spindle is adjustable for different height settings. (*Reliance Electric and Engineering Co.*)

Guiding. Another of the many functions of tooling is to guide the cutting tool so as to achieve the desired end result with regard to dimensional accuracy of the finished part.

Drill-press spindles ordinarily do not have side-thrust rigidity. Drill spindles are usually built to absorb the drill-point pressure (feed) but are essentially loose in response to sideways pressures. This is particularly true when long drills or drill extension holders are used. Consequently it is conventional practice to provide drill-guide bushings in a drill jig so as to provide the necessary degree of dimensional control.

A common example is in the use of a drill bushing to guide a drill or a reamer. A drill (or a reamer) cuts on its end rather than on its sides. Thus the engagement of the hard drill bushing with the side of the drill bit does not dull the drill and does not unduly chaff the hardened drill bushing (Fig. 2).

Cutters which cut on their cylindrical peripheries (such as end mills and toothed milling cutters) cannot be so guided because the resultant friction would be detrimental to the cutting tool edges and to the intended contact guide. Milling machines achieve the desired dimensional relation between cutter and work through the rigidity built into the machine spindle and in the strength of the machine ways along their movements.

Many drilling operations are accomplished without the use of guiding bushing by using the so-called plunge-drilling method. This is done by using a rigid milling type of spindle and holding the drill in a rigid quick-change drill holder (Fig. 3). For maximum rigidity, the drill bit is stub held and projects only as far as necessary to produce a hole of the required depth. Thus the relationship between the work and the drill is a function of the machine and work movements. Plunge drilling makes possible a variety of work in an efficient manner with simple tooling for the job (Fig. 4). Spindle construction similar to that in milling machines provides the necessary rigidity. Many tape-controlled drilling machines are built around the basic plunge-drilling method.

Plunge drilling, however, has inherent limitations that should be recognized, usually governed by the drill diameter, the depth of hole to be drilled, the tolerance of the hole location, the type of surface to be drilled, and the kind of material to be drilled. For such products as are bolted together, requiring the drilling and tapping of holes in one member and of providing bolt clearance holes in the other member, plunge drilling is usually sufficiently accurate for the dimensional needs.

Multiple tool functions. Automation, which may be defined as continuous automatic production, although not basically new, achieved a new peak of adaptation during the 1950s. This has brought about many new machine-tool and tooling problems and innovations.

The current trend in machine tools for automation is to design and produce basic machine tool

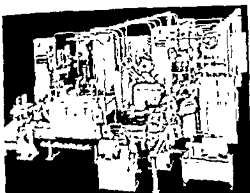


Fig 5. In this automatic transfer machine, pallets hold three parts each and shuttle through successive machine work stations Operator manually loads machine, unloading station at left is automatic. Washer automatically cleans pallets before they return to loading station (Cross Co)

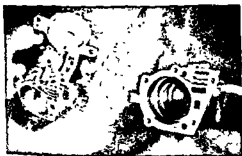


Fig 6. Two views of work piece produced in machine of Fig 5 showing results of various milling, boring, drilling, chamfering, and tapping operations (Cross Co)

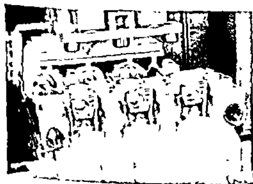


Fig 7. Work-holding pallet used in conjunction with the continuous automatic production (automation) machine (called a transfer machine) of Fig 5 (Cross Co)

building blocks, each of which is a separate machine tool, that can be linked with others (with or without special units) to form *an automated production line* that performs a series of milling, drilling, and similar machining operations (Fig 5) In this way basic machine-tool building blocks perform a variety of jobs (Fig. 6) The standard machine tool building block also offers considerable protection (through retooling or salvage) against obsolescence of an expensive automatic transfer machine

The tooling for such operations often follows the pattern that the work is located and secured to a pallet which is then automatically transferred from one operation to another (Fig. 7). The pallet has master locating surfaces so that at each work station the necessary basic alignment is reestablished and the pallet automatically secured into place. Such a procedure requires one pallet for each work station and such pallets as are required at the loading and inspection stations. *See* AUTOMATION, JIG, FIXTURE, AND DIE DESIGN; MACHINING OPERATIONS. [J.I K.]

Bibliography: *ASTE Handbook Committee, Tool Engineers' Handbook*, 2d ed, 1959

Tooth

A structure of varied type and function, present both in vertebrates and invertebrates. Teeth usually are employed in the manipulation of food; however, they may serve animals as weapons in attack or defense Among the invertebrates, teeth may be found in mollusks which possess a radula, in some annelid worms, and in many arthropods. Two types of teeth are found in vertebrates Horny teeth occur in cyclostomatous fish, many amphibian larvae, and the mammal *Ornithorhynchus*, the duck-billed platypus, while bony or true teeth are common to most other vertebrates A few vertebrates, such as certain species of whales and edentates, lack teeth

The distinguishing and universal characteristic of the teeth in vertebrated animals is dentin, a calcified tissue permeated with tubules into which extend the processes of the formative cells or odontoblasts. The teeth in the sharklike fishes are related to scales In many of the more common bony fishes, the mouth is provided with teeth which frequently fall out and are replaced. This type of *dentition is called polyphyodont.*

In the human dentition, which is heterodont, the teeth differ in shape and function. They are specialized to perform different aspects of their masticatory function—incisors cut and saw the food, cuspids and bicuspids seize and tear it, and molars grind or pulverize it

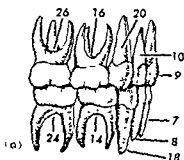
Dentition. Man is diphyodont, possessing two sets of teeth The deciduous dentition is composed of 20 teeth, the permanent teeth number 32. These replace the primary teeth and include 12 permanent molars which have no deciduous predecessors and vary considerably in their time of appearance (Fig. 1*a* and *b*)

The deciduous dentition is important because the integrity of the permanent arch depends on its care and health. A premature loss of a primary molar should not be neglected A break or irregularity in the normal contacts between the teeth caused by the loss or extraction of a single tooth may result in their abnormal positioning (malocclusion). The teeth of opposing arches interdigitate so that the

upper arch overlaps the lower one, and each tooth is opposed by two teeth of the other arch except for the third molars and lower central incisors.

Structure. Anatomically, each tooth consists of a crown and a root (Fig. 2). The root is embedded within a bony socket of the jaw and supports the crown in its masticatory function. The body of the tooth consists of dentin which is covered by enamel in the crown and by cementum in the root. The central cavity of the tooth is filled with soft connective tissue, the pulp, which forms and nourishes the dentin. Nerves and blood vessels enter through the tip of the root. The human tooth is not fused with the bone but suspended within the socket by means of the fibers of the periodontal membrane. The latter is a ligament which attaches the cementum and the tooth to the alveolar bone. This type of fixation is termed gomphosis. The gingiva is the part of the oral mucosa which covers the alveolar bone and the neck of the tooth (Fig. 2).

The enamel serves to resist abrasive wear and consists of prisms and interprismatic substances. It is the hardest tissue in the body and is composed almost entirely (97%) of inorganic salts, mainly calcium phosphate. Enamel stands out as the tissue which is entirely avascular and acellular; once lost, it cannot be replaced except by artificial means.



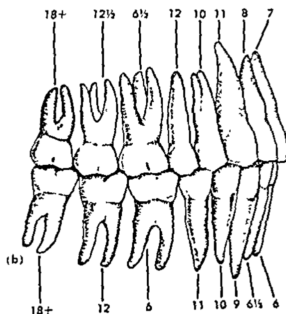


Fig. 1. (a) Eruption time of deciduous teeth in months. (b) Eruption time of permanent teeth in years. (*Adapted from M. Massler and I. Schour, Atlas of the Mouth, plates 2 and 3, American Dental Assoc., 1952*)

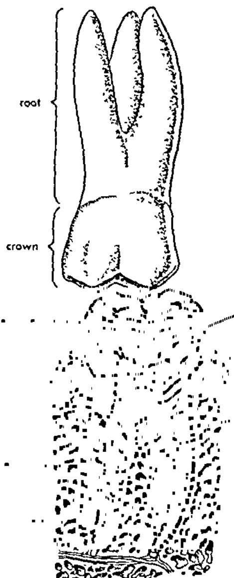


Fig 2 The structure of a molar tooth and its surrounding tissues. (*Adapted from I Schour, How teeth grow, J. Am. Dental Assoc., 30 133, fig. 1, 1943*)

The dentin gives to the tooth its general form and elastic strength. It consists of about 67% inorganic salts, mainly calcium phosphate, the rest being organic material and water. Since the organic matrix is laid down first and then calcified, there is normally a narrow border of uncalcified dentin called predentin next to the pulpal surface. In rickets or calcium deficiency, the predentin layer becomes much wider. The pulp reacts to cutting or wear of the dentin and exposure of its tubules by the formation of secondary reparative dentin.

The cementum covers the surface of the root. It is a modified bone which grows slowly but continuously. It is relatively thin and serves to attach the periodontal fibers which suspend the tooth within its socket. In case of injury, the cementum can regenerate and reattach the suspensory fibers to the tooth.

The periodontal membrane is the fibrous connective tissue which fills the space between the root of the tooth and bony wall of its alveolus socket. It has three functions: to connect the tooth with the adjacent hard and soft tissues, to form bone on the

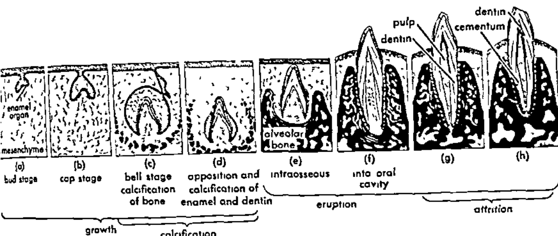


Fig 3 (a-h) Life cycle of tooth. (*Adapted from I Schour and M. Massler, Studies in tooth development, J. Am Dental Assoc., 27 1778, fig. 1, 1940*)

wall of the socket and cementum on the root, and to serve as the seat of the sense of touch for the tooth. The arrangement of the principal fibers is beautifully adapted to sustain the tooth against the masticatory forces and to help it to absorb shock.

The alveolar bone is that part of the jawbone which forms the sockets of the teeth. The sockets persist only as long as the tooth is present and functional. It is highly adaptable to tooth function

for the loss of teeth in the younger age groups, after 35 years of age the greatest single cause of

deum. At 20 points of the dental lamina corresponding to the future position of the primary teeth, epithelial buds grow into the underlying connective tissue and develop into the organs which form the enamel and guide the subjacent connective tissue to form the dental papilla, or the future dentin forming organ and pulp

The tooth, during its life cycle, develops in an orderly manner through a number of stages (Fig 3) Initiation and proliferation of the odontogenic cells is followed by their differentiation into enamel- and dentin forming cells and by their morphodifferentiation or arrangement to shape the future tooth.

Deposition of hard tissue begins at the cuspal tips or growth center of the crown It proceeds in layers along an *incremental* pattern as illustrated in Fig. 3 and is closely followed by calcification. Disturbances in calcification can occur as a result of relatively mild systemic disturbances involving calcium metabolism The calcifying dental tissues, therefore, serve as biologic recorders which reflect the health of the growing individual. Thus the birth experience with its profound neonatal adjustments produces an accentuated incremental layer or ring within the enamel and dentin at the level of tissue formed at the time of birth. These are called the birth or neonatal rings and are found in all deciduous teeth Such rings constitute the basis for tooth ring analysis, a technique analogous to tree ring analysis.

Nutritional deficiencies in minerals or vitamins may lead to disturbances in the formation and calcification of the teeth. Excessive fluorides in the water supply can produce discoloration and poor calcification of the teeth. Fortunately, the amount of fluorides (one part of fluorine per million parts of water) which gives the enamel protection against dental decay is harmless and does not disturb tooth development.

The piercing of the tooth through the oral mucosa is only a momentary and transitory event because eruption continues throughout the life of the tooth. Eruption is rapid until the tooth meets its antagonist, it then proceeds very slowly to compensate for the normal attrition of the tooth surface *See* DENTITION. [I S.]

Bibliography· M. Massler and I Schour, *Atlas of the Mouth*, 1952, B A. Willier, P. A. Weiss, and V Hamburger (eds.), *Analysis of Development*, 1955.

Tooth disorders

Disturbances of the teeth and related structures Tooth disorders often produce marked impairment of general health since mastication and subsequent digestion may be reduced In addition, bacteria or their products may invade other tissues from a dental infection, either by direct extension or by passing into the blood stream There is a strong positive

relationship between general bodily health and that of the oral cavity and teeth. *See* TOOTH.

Dentine and enamel, despite their original resistance, show little reactive or reparative capacity against injury or infection.

Developmental defects include hereditary failures in formation or in the deposition of intact enamel. Symmetrical defects of enamel and dentine are often associated with faulty phosphorus and calcium metabolism, as in rickets and tetany. In congenital syphilis, a typical defect occurs that is known as Hutchinson's teeth. These are deformed and characteristically protrude at the tooth corners while showing indentation in the center, in contrast to normal teeth that are slightly rounded toward the biting surface. *See* SYPHILIS.

Dental caries, or cavity formation, results from the gradual deterioration of the enamel, dentine, and finally the tooth pulp. It is due to many contributing factors or combinations and no one explanation is satisfactory The addition of fluoride compounds to drinking water has met with some success in the reduction of cavities, but the process is still controversial Fluorides act by combining with calcium salts, thereby increasing the resistance of the outer enamel to cavitation. Excessive fluorides will, however, cause enamel discoloration

Dentoalveolar abscesses are acute or chronic inflammations from bacterial infections that occur in the root canals of the jaws Occasionally a gum boil may be present but more often only x-ray examination will reveal the actual pocket of infection

Periodontal disease, or pyorrhea, is an inflammation of the gum margin and tooth sheath (periosteum) from local irritation or infection. *See* PYORRHEA.

Toothache, or ondontalgia, is pain arising from stimulation of the dental nerves by any process such as inflammation It may occasionally occur as a reflex stimulation of the nerve roots and fibers from a lesion between the teeth and the brain.

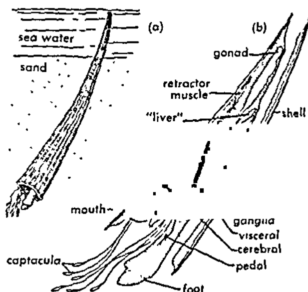
Diseases of the jaws are intimately related to tooth disorders since the upper and lower jaws form series of pockets for the teeth and carry their blood and nerve supply The most common jaw disorders fall into four categories: (1) the inflammations of the jaw bones, such as osteomyelitis, (2) cysts of the jaws, both related to the teeth and separate from them, (3) tumors of benign or malignant nature, such as osteomas and multiple myeloma; and (4) involvement of the jaws in systemic disease. Examples of the last would include certain rare bone disorders, generalized skeletal disturbances, and those produced by endocrine dysfunction. [E.G.ST.]

Tooth shell

A member of the class Scaphopoda, phylum Mollusca. There are approximately 200 living species. About 300 fossil species are known, dating from the Devonian. Tooth shells are sometimes called elephant tusk shells.

Tooth shells are of no particular significance either economically or medically, other than their interesting shape. They have the shell and mantle modified into a slender, slightly curved and tapered tube, open at both ends. They are all marine and are usually found in deep water, although they may occur at all depths up to 15,000 ft. The pointed foot extends from the larger end of the tube. They live foot downward, buried obliquely in the mud or sand.

Food, consisting of minute plants and animals, is captured by retractile, sensory tentacles called captacula, covered with cilia. Several of these tentacles extend around the mouth from the larger end of the shell. There is no head. Respiration is



The tooth shell, *Dentalium* (a) Position in life. (b) Internal structure from the left side, diagrammatic. *(From T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957)*

accomplished by the mantle Otherwise their general structure is similar to that of the snails

The sexes are separate The young undergo a short larval period before sinking to the bottom.

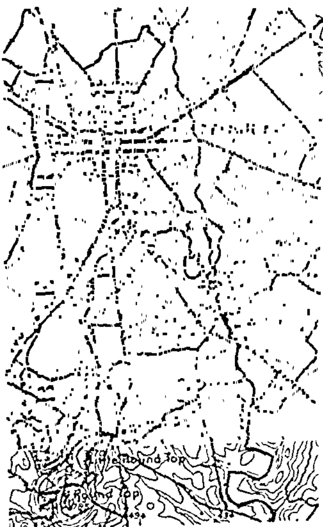
Pacific Coast Indians formerly used shells of *Dentalium pretiosum* for money and ornaments. Other than those washed ashore, they were obtained by dredging from canoes with long rakes This species occurs on sandy beaches from California to Alaska and from the low tide mark to considerable depths *See* SCAPHOPODA, SNAIL. [J.D.R.]

Topaz

A mineral best known for its use as a gem stone Crystals are usually colorless but may be red, yellow, green, blue, or brown The wine yellow variety is the one usually cut and most highly prized as a gem. Corundum of similar color sometimes goes under the name of Oriental topaz. Citrine, a yellow variety of quartz, is the most common substitute and may be sold as quartz topaz. *See* GEM

Topaz is a nesosilicate with chemical composition $Al_2SiO_4(F,OH)_2$. The mineral crystallizes in the orthorhombic system and is commonly found in well developed prismatic crystals with pyramidal terminations. It has a perfect basal cleavage

Substantial areas and difficult terrain are surveyed with less cost by aerial surveys (*see* PHOTOGRAMMETRY). Smaller areas, requiring but one or a few days of field-party time, may be surveyed and mapped for less money than the aerial-survey mobilization costs of flying and photo processing. Visibility of the ground and other features to be mapped also influence the choice. Foliage and snow may conceal topography from the aerial camera, in many areas limiting the times of year when photogrammetry can be used effectively. The degree of accuracy required for elevations also may rule out aerial surveys. Theoretically, any precision is available, but the difficulty of maintaining flight lines at low altitudes, together with surface variables like grass height, casts doubt on the accuracy of photogrammetric contour intervals of less than 2 ft. Above this practical limit, there is no difference between the accuracies of ground-surveyed and aerial-surveyed topography. Both usually must comply with the same generally accepted standards of mapping accuracy. These are that (1) 90% of all elevations determined from the contours must be correct within one-half the contour interval; the remainder must be correct within one contour interval, and (2) 90% of all planimetric features must be plotted within one-fortieth of an inch of their true coordinate positions and the remainder must be plotted within 1/30 in.



Topographic map, scale 1.62,500, contour interval 20 ft. Spot elevations shown at road intersections. (USGS)

Ground surveying. Topographic surveying on the ground entails measurement of the horizontal positions and elevations of enough points to describe the terrain and all features to be shown on the map. Measurements are taken from control points, usually by transit and tape or stadia, or by plane table and stadia. Slope measurements, reduced trigonometrically to horizontal and vertical distances, are of sufficient precision to accomplish standard map accuracies for most scales and contour intervals. *See* SURVEYING.

The choice of points to be observed is simple for features such as buildings, fence corners, and bends in streams; it becomes complex where the configuration of irregular terrain is to be described. The rodman or tapeman must give the instrument man sightings on all significant changes of slope and contour direction without wasting time on intervening points.

Transit survey. In a transit survey the instrument is set up at a control point and oriented. Directions to all observed points are recorded, and where distance measurements are feasible, the horizontal distance and elevation are computed and recorded. Where directions only are observed, the points must be sighted for direction from a subsequent control point to fix their horizontal positions. The elevations of these points may be obtained by auxiliary means, such as hand-leveling. The transitman or a notekeeper must make his notes as descriptive as possible. Quantitative data and word descriptions are often supplemented by sketches, including contours.

In the office, transit-survey field data are plotted on compilation sheets. Points on contour lines are established by proportional interpolation between observed elevations, and the contour lines are drawn to connect the points interpolated for each contour line

Where high precision is required, as for earthwork for a relatively flat area, it can be staked in gridiron pattern so the elevations of all line intersections can be observed by differential leveling techniques.

Plane table survey Map compilation is completed in the field with the plane table (*see* ALIDADE, PLANE TABLE) Where the control system has been established, its points are plotted on the plane-table sheet, and the plane table is set up and oriented at a control point The alidade straight edge is set on the corresponding plotted point and a series of stadia sightings are made on chosen terrain points Scale distances establish the plotting of each terrain point on the plane-table sheet, and its elevation is noted alongside Contours and planimetric lines are drawn as the survey progresses

Aerial surveying. Aerial topographic surveying brings the terrain into the photogrammetry laboratory Two or more overlapping photos are projected stereoscopically to establish a three-dimensional image—literally a visual scale model of the

terrain—on which an instrument operator can measure the horizontal distance and elevation difference between any two points he can see.

Photographing. The choice of airplane flight altitude depends on the degree of vertical accuracy required for the map and on the precision of the stereoplotting instrument. Based on the cited standards of map accuracy, the prescribed contour interval indicates the degree of vertical accuracy required. Where a cartographic camera with 6-in. focal length takes pictures for use in a stereoplotting instrument of ordinary precision, the flight altitude in feet should be 800 to 1000 times the contour interval in feet to assure standard map accuracy Called the C factor, this multiplier may be 1200 or more for highly precise stereoplotting instruments.

Photographs are taken straight down The frequency of exposure is timed to provide at least 50% and usually 60% of picture overlap in the direction of the flight line. If the single flight line does not give sufficient breadth of coverage, additional parallel lines are flown with side-overlap of 30% or more.

Plotting. In the stereoplotting instrument, planimetry and contours are drawn on a compilation sheet, corresponding with the plane-table sheet, called the manuscript The sheet should be a dimensionally stable material, so the accuracy of stereoplotting can be preserved for map drafting

Map drafting. Photogrammetric map drafting, as well as plane-table map drafting, consists of tracing the manuscript lines and inscribing symbols and legends on the tracing in an orderly manner The concept of order-of-drafting priority is important: lines are drawn first in their correct positions; symbols locating objects also must be in their correct positions; legends identifying small features (and therefore having to be adjacent to those features) come next; then more general information, such as area and ownership of land plots, or names of political subdivisions, are inscribed in the remaining spaces. A north arrow, identified as true north, magnetic north or some other reference system, is essential The coordinate grid is drafted precisely, and a graphic scale is provided as a precaution against subsequent shrinkage and expansion of the tracing and any prints made from it. In addition, there is a title naming the map and stating ownership, authorship and other pertinent facts A legend of special symbols may be included. One color (usually black) or several colors of ink may be used for the drafting Reproduction may be by any of several methods, ranging from blue-printing to lithography *See* CARTOGRAPHY; MAP PROJECTIONS [R H DO]

Bibliography: American Society of Photogrammetry, *Manual of Photogrammetry*, 2d ed., 1952

Topology

The study of topological spaces and continuous maps. Using elementary set theory, precise mathematical definitions of topological space and continuous map will be given, and then these will be illustrated by examples.

Sets. A set is any collection of things or objects. If X is a set and x is a member of the set X, one writes $x \in X$, which reads in words x belongs to X. The symbol $\in$ denotes membership. An element or member of a set X is also called a point of X. The rational numbers form a set frequently denoted by Q, and $x \in Q$ means that x is a rational number.

If X and Y are sets, then $X \cup Y$ is the set of elements which belong to either X or Y (including those elements which belong to both X and Y). The notation $X \cup Y$ is read X union Y. Similarly if X and Y are sets, then $X \cap Y$ is the set of elements which belong to both X and Y, and the notation $X \cap Y$ is read X intersect Y.

Suppose for example, that Q is the set of rational numbers, X is the set of rational numbers greater than or equal to zero, and Y is the set of rational numbers less than or equal to zero. Under these conditions $X \cup Y = Q$, and $X \cap Y$ is the set whose only element is the number zero.

Suppose that for each element ι of a set I there is given a set A_ι, the set consisting of all the sets A_ι is denoted by $\{A_\iota\}_{\iota \in I}$, and is said to be a collection of sets indexed on the set I If $\{A_\iota\}_{\iota \in I}$ is a collection of sets indexed on the set I, then $\bigcup_{\iota \in I} A_\iota$ is the set consisting of those elements which belong to at least one of the sets A_ι, and is called the union of $\{A_\iota\}_{\iota \in I}$. Similarly $\bigcap_{\iota \in I} A_\iota$ is the set consisting of those elements which belong to every one of the sets A_ι, and is called the intersection of $\{A_\iota\}_{\iota \in I}$ In case I is the set with exactly two elements 1 and 2, $\{A_\iota\}_{\iota \in I}$ has two members A_1 and A_2. Moreover, in this case $\bigcup_{\iota \in I} A_\iota = A_1 \cup A_2$, and $\bigcap_{\iota \in I} A_\iota = A_1 \cap A_2$.

If X is a set, then a set A is a subset of X if every element of A is also an element of X For example, if X is the set of automobiles built during any one year, and A is the set of green automobiles built during that year, then A is a subset of X. Every set has a subset called the empty set. The empty set is the set with no elements whatsoever, and is frequently denoted by ϕ. The notation $A \cap B = \phi$ means that the intersection of the set A and B is empty, in other words that the sets A and B are disjoint.

Topological space. This is a set of points X, together with a collection of subsets of X called open subsets of X where the following assumptions are made:

1 ϕ and X are open subsets of X.

2. If A and B are open subsets of X, then $A \cap B$ is an open subset of X

3. If $\{A_\iota\}_{\iota \in I}$ is a collection of open subsets of X, then $\bigcup_{\iota \in I} A_\iota$ is an open subset of X

Suppose that R is the set of real numbers If $r \in R$, in other words if r is a real number, let $|r|$ denote the absolute value of r This means that if r is greater than or equal to zero, then $|r| = r$, but if r is less than zero then $|r| = -r$. In order to make the real numbers

Substantial areas and difficult terrain are surveyed with less cost by aerial surveys (*see* PHOTOGRAMMETRY). Smaller areas, requiring but one or a few days of field-party time, may be surveyed and mapped for less money than the aerial-survey mobilization costs of flying and photo processing. Visibility of the ground and other features to be mapped also influence the choice. Foliage and snow may conceal topography from the aerial camera, in many areas limiting the times of year when photogrammetry can be used effectively. The degree of accuracy required for elevations also may rule out aerial surveys. Theoretically, any precision is available, but the difficulty of maintaining flight lines at low altitudes, together with surface *variables like grass height, casts doubt on the* accuracy of photogrammetric contour intervals of less than 2 ft. Above this practical limit, there is no difference between the accuracies of ground-surveyed and aerial-surveyed topography. Both usually must comply with the same generally accepted standards of mapping accuracy. These are that (1) 90% of all elevations determined from the contours must be correct within one-half the contour interval; the remainder must be correct within one contour interval, and (2) 90% of all planimetric features must be plotted within one-fortieth of an inch of their true coordinate positions and the remainder must be plotted within 1/20 in.



Topographic map, scale 1:62,500, contour interval 20 ft. Spot elevations shown at road intersections. (USGS)

Ground surveying. Topographic surveying on the ground entails measurement of the horizontal positions and elevations of enough points to describe the terrain and all features to be shown on the map. Measurements are taken from control points, usually by transit and tape or stadia, or by plane table and stadia. Slope measurements, reduced trigonometrically to horizontal and vertical distances, are of sufficient precision to accomplish standard map accuracies for most scales and contour intervals. *See* SURVEYING.

The choice of points to be observed is simple for features such as buildings, fence corners, and bends in streams; it becomes complex where the configuration of irregular terrain is to be described. The rodman or tapeman must give the instrument man sightings on all significant changes of slope and contour direction without wasting time on intervening points.

Transit survey. In a transit survey the instrument is set up at a control point and oriented. Directions to all observed points are recorded, and where distance measurements are feasible, the horizontal distance and elevation are computed and recorded. Where directions only are observed, the points must be sighted for direction from a subsequent control point to fix their horizontal positions. The elevations of these points may be obtained by auxiliary means, such as hand-leveling. The transitman or a notekeeper must make his notes as descriptive as possible. Quantitative data and word descriptions are often supplemented by sketches, including contours.

In the office, transit-survey field data are plotted on compilation sheets. Points on contour lines are established by proportional interpolation between observed elevations, and the contour lines are drawn to connect the points interpolated for each contour line.

Where high precision is required, as for earthwork for a relatively flat area, it can be staked in gridiron pattern so the elevations of all line intersections can be observed by differential leveling techniques.

Plane table survey. Map compilation is completed in the field with the plane table (*see* ALIDADE; PLANE TABLE) Where the control system has been established, its points are plotted on the plane-table sheet, and the plane table is set up and oriented at a control point. The alidade straightedge is set on the corresponding plotted point and a series of stadia sightings are made on chosen terrain points Scale distances establish the plotting of each terrain point on the plane-table sheet, and its elevation is noted alongside. Contours and planimetric lines are drawn as the survey progresses.

Aerial surveying. Aerial topographic surveying brings the terrain into the photogrammetry laboratory. Two or more overlapping photos are projected stereoscopically to establish a three-dimensional image—literally a visual scale model of the

subset W of Y such that $V \times W$ is a subset of U. When $X \times Y$ is considered as a topological space with open subset defined in this manner, it is called the product space of X and Y.

Having defined the concept of product space it is now easy to define that of euclidean n-space. However, before doing this it is well to remark that if X, Y, and Z are topological spaces, then the space $(X \times Y) \times Z$ and $X \times (Y \times Z)$ are the same. In other words the operation of taking the product space of several topological spaces is associative. Now euclidean 1-space is just the space of real numbers R. Euclidean 1 space also called the real line or just the [illegible]

[illegible] by R^n, euclidean $(n+1)$ space is defined to be $R^n \times R$ and denoted by R^{n+1}.

The preceding inductive definition of R^n is convenient for some purposes, but it is useful to have a direct description also. This may be accomplished by first letting R^n denote the set of n-tuples

$$(x_1, \ldots, x_n)$$

such that each x_i is a real number, and then defining a metric on R^n by letting the distance

$$\rho((x_1, \ldots, x_n), (y_1, \ldots, y_n))$$

from the point $(x_1, \ldots, x_n)$ to the point

$$(y_1, \ldots, y_n)$$

be the square root of $\Sigma_{i=1}^{n}(x_i - y_i)^2$. Thus if $n = 2$, then $R^2 = R \times R$ and the distance $\rho((x_1,x_2),(y_1,y_2))$ from the point (x_1,x_2) to the point (y_1,y_2) is the square root of $(x_1 - y_1)^2 + (x_2 - y_2)^2$. The space R^2 is frequently called the plane or the euclidean plane and corresponds to the intuitive notion of a flat or plane surface. The space R^3 is sometimes called just space and corresponds to the intuitive idea of space. The distance between two points (x_1,x_2,x_3) and (y_1,y_2,y_3) in R^3 is the square root of $(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2$.

The n-dimensional sphere, denoted by S^n, is the subspace of R^{n+1} consisting of those points whose [illegible]

[illegible] $(0,0)$ is exactly 1 Consequently the circle is the [illegible]

[illegible] equal to 1 It is pictured as the surface of an ordinary solid ball such as a croquet ball The 2 sphere is an example of the general notion of surface There are many other examples of surfaces, a common one being the torus which is just the topological space $S^1 \times S^1$. It may be thought of as the surface of a doughnut or as the inner tube of an automobile tire

Two topological spaces which are homeomorphic cannot be distinguished by the methods of topology Any topological property of one is also a topological property [illegible] calls any [illegible] sphere S^n, [illegible] any circle in the plane is homeomorphic to the standard circle S^1. Suppose that X is the topological space which is the boundary of a square in the plane. Let Y be a circle in the plane which completely contains X in its interior. Choose a point * in the plane inside the square. Now for every point $x \in X$, let $f(x)$ be the point of Y obtained by proceeding in a straight line from * to x and then continuing along the same straight line from x to the point $f(x)$ of Y. The function $f: X \rightarrow Y$ thus obtained is a homeomorphism. Consequently X, the boundary of a square, is homeomorphic to a circle. Similarly it may be proved that a square in the plane is homeomorphic to the disk bounded by a circle.

Since homeomorphic spaces cannot be distinguished by topological methods, one of the important problems of topology is to determine whether two topological spaces are homeomorphic or not. For example, is the sphere S^1 homeomorphic to the sphere S^2? Intuitively it does not seem so, but one would like a proof. One proof may be obtained by giving topological significance to the idea of dimension

Suppose that X is a topological space. An open covering of X is a collection of sets $\{A_i\}_{i \in I}$ such that each A_i is an open subset of X, and $X = \cup_{i \in I} A_i$. If $\{A_i\}_{i \in I}$ and $\{B_j\}_{j \in J}$ are open coverings of X, the covering $\{A_i\}_{i \in I}$ is said to be finer than the covering $\{B_j\}_{j \in J}$ if for each index $i \in I$ there exists an index $j \in J$ such that A_i is a subset of B_j. Let $\{A_i\}_{i \in I}$ be an open covering of the space X This covering of X is said to have dimension less than or equal to n if the intersection $A_{i_0} \cap \quad \cap A_{i_n}$ of any $(n+1)$ distinct sets of the covering is empty Thus an open covering $\{A_i\}_{i \in I}$ has dimension less than or equal to zero if $A_i \cap A_j$ is empty for A_i different from A_j. It has dimension less than or equal to 1 if for every three sets A_{i_1}, A_{i_2}, A_{i_3} such that no two are equal the intersection $A_{i_1} \cap A_{i_2} \cap A_{i_3}$ is empty, and so forth The topological space X has dimension n if n is the least integer with the property that for every open covering of X there is a finer open covering of X which has dimension less than or equal to n. From the definition of dimension just given it is clear that if X is an n-dimensional topological space and Y is homeomorphic with X, then Y is an n-dimensional topological space. In other words the notion of dimension is a topological invariant. Moreover, it is possible to prove that euclidean n-space R^n is an n-dimensional topological space, and that the n-sphere S^n is an n-dimensional topological space. This means among other things that the circle S^1 is not homeomorphic to the 2-sphere S^2. Further it means that there is topological significance to the n of S^n or R^n.

There are other important properties of topological spaces which are defined by using the notion of open covering. Probably the most important of these is compactness. A topological space X is compact if and only if for any open covering $\{A_i\}_{i \in I}$ of X there is a

ε_x greater than zero having the property that if for some real number y, $|y - x| < \varepsilon_x$ then $y \in U$. In other words if $x \in U$ then any real number sufficiently close to x belongs to U also. With this definition of open subset of the real numbers it is not difficult to verify that the axioms for a topological space are satisfied. When one talks of the real numbers in mathematics, one usually means the real numbers as a topological space, that is to say the set of real numbers together with the collection of open subsets that are defined above.

Continuous map. In order to continue the discussion of topology, it is necessary to introduce some further notions of set theory. First suppose that X and Y are sets. Define the product of the sets X and Y to be the set consisting of pairs of elements (x,y) such that $x \in X$ and $y \in Y$. The product of X and Y is denoted by $X \times Y$. Now a function f from X to Y is defined to be a subset f of the product $X \times Y$ such that if $x \in X$ there exists a unique $y \in Y$ such that $(x,y) \in f$. In this case y is said to be the value of f at x, and is denoted by $f(x)$. Intuitively a function from X to Y is thought of as a rule which assigns to each element of the set X an element of the set Y. The standard mathematical notation for a function f from X to Y is $f:X \rightarrow Y$.

Let $f:X \rightarrow Y$ be a function and suppose U is a subset of Y, then $f^{-1}(U)$ is the subset of X consisting of those points x such that $f(x) \in U$ If X and Y are topological spaces, then $f:X \rightarrow Y$ is a continuous map, or continuous function, (1) if f is a function from X to Y, and (2) if, whenever U is an open subset of Y, the set $f^{-1}(U)$ is an open subset of X If X is any topological space, then the identity function $\iota:X \rightarrow X$ defined by $\iota(x) = x$ is a continuous map called the identity map of X. Suppose X, Y, and Z are topological spaces, $f: X \rightarrow Y$ is a continuous map, and $g:Y \rightarrow Z$ is a continuous map In this case the function $g \cdot f:X \rightarrow Z$ defined by $(g \cdot f)(x) = g(f(x))$ is a continuous map.

The topological spaces X and Y are said to be homeomorphic if there exist continuous maps $f: X \rightarrow Y$ and $g:Y \rightarrow X$ such that $f \cdot g: Y \rightarrow Y$ is the identity map of Y and $g \cdot f: X \rightarrow X$ is the identity map of X In this case the maps f and g are said to be homeomorphisms, and g is frequently denoted by f^{-1} The symbol f^{-1} is read f inverse.

Again assume that R is the topological space of real numbers. Define a function $f:R \rightarrow R$ by $f(x) = 0$ if x is less than zero, and $f(x) = 1$ if x is greater than or equal to zero. The function $f:R \rightarrow R$ is not a continuous map. This is because $f(0) = 1$, but there are points x arbitrarily close to 0 such that $f(x) = 0$. In other words the function f is not smooth; it jumps at zero. Intuitively a continuous map $f:X \rightarrow Y$ is a function such that if x and x' are close together then $f(x)$ and $f(x')$ are also close together where the notion of closeness is determined by the open subsets of X and Y respectively.

Metrics. When the topology on the real numbers was defined, this was a special case of defining a topology by using a metric. Let X be a set. A metric on X is a function $\rho:X \times X \rightarrow R$ such that the following axioms obtain:

1. If x and x' belong to X, then $\rho(x,x')$ is greater than or equal to zero, and $\rho(x,x') = 0$ if and only if $x = x'$.
2. If x and x' belong to X, then $\rho(x,x') = \rho(x',x)$.
3. If x, x', and x'' belong to X, then $\rho(x,x'')$ is less than, or equal to, $\rho(x,x') + \rho(x',x'')$.

If $\rho:X \times X \rightarrow X$ is a metric, then $\rho(x,x')$ is called the distance from x to x'. Axiom 1 says that the distance between any two points of X is greater than or equal to zero, and is different from zero if the points are different. Axiom 2 says that the distance from x to x' is the same as the distance from x' to x. Axiom 3, the so-called triangle axiom, says intuitively that it is shorter to proceed from x to x'' along a straight line than it is first to proceed from x to x' along a straight line and then proceed from x' to x'' along another straight line.

If X is a set, and $\rho:X \times X \rightarrow R$ is a metric on X, one says that a subset U of X is open if, whenever $x \in U$, there exists a real number ε_x greater than zero such that if x' is another point of X and $\rho(x,x')$ is less than ε_x then x' also belongs to U. In other words if $x \in U$ and x' is a point which is very close to x, then x' also belongs to U. One verifies easily that if open subset of X is defined as above, then axioms for the open subsets of a topological space are verified. Thus any set X with a metric ρ $X \times X \rightarrow R$ defines a topological space.

When R is the set of real numbers and $\rho:R \times R \rightarrow R$ is defined by $\rho(x,x') = |x - x'|$ it may be proved that ρ is a metric on R Then clearly the notion of open subset defined from this metric is exactly the notion of open subset defined earlier.

Suppose that X is a set, $\rho: X \times X \rightarrow X$ is a metric on X, Y is a set, and ρ' $Y \times Y \rightarrow Y$ is a metric on Y. Further assume that the notion of open subset of X is defined by using the metric ρ, and that the notion of open subset of Y is defined by using the metric ρ'. In this case it may be proved that $f: X \rightarrow Y$ is a continuous map if and only if for every $x \in X$ and ε_x greater than zero, there exists δ_x greater than zero such that if $\rho(x,x')$ is less than δ_x then $\rho'(f(x),f(x'))$ is less than ε_x. Thus the idea that f $X \rightarrow Y$ is continuous whenever x and x' are close together implies that $f(x)$ and $f(x')$ are close together; the idea is precise when the open subsets of X and Y are defined by metrics.

Construction of topological spaces. One of the most important processes in topology is the construction of new topological spaces from old ones Two particularly common and useful methods of doing this will be illustrated here. First suppose X is a topological space and A is a subset of X. A subset U of A is open if there exists an open subset V of X such that $A \cap V = U$ When A is considered as a topological space with the notion of open subset de-

U there exists an open subset V of X and an open

is not compact and has the same homotopy type as R^n. The fact that at present the preceding result cannot be replaced by one saying that under the same conditions X is either homeomorphic with S^n or R^n unless n is less than or equal to 2, seems to be one of the main obstacles to proving classification theorems for manifolds of dimension greater than 2.

In the last fifty years the study of topology has changed considerably. Instead of attacking problems directly, algebraic invariants are attached to topological spaces, and then these invariants are studied. The main impetus to starting to work in this direction was given by the French mathematician Henri Poincaré about fifty or sixty years ago. The work of Poincaré was done chiefly in connection with the problem of classifying manifolds. He conjectured that if X is a compact connected 3-manifold and every map $f: S^1 \to X$ is homotopically trivial, then X is homeomorphic with a 3 sphere. As was mentioned earlier there is as yet no proof of this result, though it still seems possible that one may be found.

The study of topology by means of the algebraic invariants of topological spaces is in reality a comparatively new field of mathematics. Since most of these invariants are rather difficult to describe, no attempt to do so will be made here. The definition of just one set of these invariants, the homotopy groups will be outlined. These invariants are the easiest to [illegible] points x and x' belonging to X, there is a continuous map $f: I \to X$ such that $f(0) = x$ and $f(1) = x'$, in other words if one can draw an arc between any two points of the space. A manifold is connected if, and only if, it is pathwise connected.

Suppose that X and Y are topological spaces, $x_0 \in X$ and $y_0 \in Y$ are chosen and called base points. A map $f: X \to Y$ is said to preserve base points if $f(x_0) = y_0$. Two maps $f: X \to Y$ and $g: X \to Y$ which [illegible] to the [illegible] $\times X \to$ [illegible]

1. $F(0,x) = f(x)$ for any $x \in X$.
2. $F(1,x) = g(x)$ for any $x \in X$.
3. $F(t,x_0) = y_0$ for any $t \in I$.

Now one says that two maps $f, g: X \to Y$ which preserve base point are equivalent if they are homotopic relative to the base points. The set of such equivalence classes of maps is denoted by $\pi((X,x_0),(Y,y_0))$. Such an equivalence class is called a homotopy class of maps.

For the sphere S^n choose a base point $e_n \in S^n$ once and for all. Let $\pi_n(X,x_0)$ be $\pi((S^n,e_n),(X,x_0))$ for every space X with base point x_0. Let $S^n \vee S^n$ be the space obtained by taking two copies of S^n and identifying the point e_n in one copy with the point e_n in the other copy. This space may be thought of as two tangent n-spheres. Define a map $\theta: S^n \to S^n \vee S^n$ by collapsing an equator of S^n through the base point to obtain two tangent spheres. Suppose that $f, g: S^n \to X$ are maps which preserve base points. Define $f \vee g: S^n \vee S^n \to X$ by mapping points of the first tangent sphere by means of f and of the second by means of g. This definition is legitimate since $f(e_n) = x_0 = g(e_n)$. Now $(f \vee g) \cdot \theta: S^n \to X$ and preserves base point. One verifies that if f' is homotopic to f and g' is homotopic to g, then $(f' \vee g') \cdot \theta$ is homotopic to $(f \vee g) \cdot \theta$ where all homotopies are relative to the base points. If for any such map f, one denotes the homotopy class of f by $[f]$, then $[(f \vee g) \cdot \theta]$ depends only on $[f]$ and $[g]$. Then mapping $\varphi: \pi_n(X,x_0) \times \pi_n(X,x_0) \to \pi_n(X,x_0)$ defined by

$$\varphi([f],[g]) = [(f \vee g) \cdot \theta]$$

determines a group operation in the set $\pi_n(X,x_0)$ which is now called the n-dimensional homotopy group of X relative to the base point x_0.

If X is pathwise connected, then the group $\pi_n(X,x_0)$ is independent of the choice of base point x_0. Further if X and Y are pathwise connected spaces having the same homotopy type, then for any $x_0 \in X$ and any $y_0 \in Y$ the groups $\pi_n(X,x_0)$ and $\pi_n(Y,y_0)$ are isomorphic.

The group $\pi_1(X,x_0)$ is also called the fundamental [illegible] some thirty years later by Witold Hurewicz. It is not difficult to prove that if X is a space having the homotopy type of a point, then $\pi_n(X,x_0)$ has a single element. Further it may be proved that the group $\pi_q(S^n,e_n)$ has a single element if q is less than n, but is isomorphic with the group of integers if $q = n$. This proves that S^n does not have the homotopy of a point.

Even though the groups $\pi_q(X,x_0)$ are abelian for $q > 1$, they are difficult to compute. It may be shown that if n is greater than 1, then the group $\pi_q(S^n,e_n)$ has more than one element for an infinite number of integers q. [J.C.MO.]

Bibliography: S. Lefschetz, *Introduction to Topology*, 1949; N. Steenrod and S. Eilenberg, *Foundations of Algebraic Topology*, 1952; R. L. Wilder, *Topology of Manifolds*, 1949.

Torbanite

A variety of coal that resembles a carbonaceous shale in outward appearance. It is fine-grained, brown to black, tough, and breaks with a conchoidal or subconchoidal fracture. The name torbanite is derived from the initial discovery site of the material in 1850 at Torbane Hill, Linlithgowshire, Scotland. Torbanite is synonymous with boghead coal and is related to cannel coal. Torbanite is derived from colonial algae identified with the modern species of *Botryococcus braunii* Kütz and its antecedent forms.

Major deposits of torbanite occur in Australia, Tasmania, New Zealand, Scotland, and South Africa. The South African deposit, which is in the Ermelo district of the Transvaal, yields from 20–100 gal of oil per ton on retorting. High-assay torbanite yields paraffinic oil, whereas low-assay material yields asphaltic oil. See COAL; SAPROPEL.

[I.A.B.]

finite number of indices $i_1, \ldots, i_n \in I$ such that $X = \bigcup_{j=1}^{n} A_{i_j}$. For any positive integer n the sphere S^n is a compact topological space. A general theorem asserts that if X and Y are compact topological spaces, then the product space $X \times Y$ is also compact. This implies that the torus is a compact topological space for it is the product space $S^1 \times S^1$ and S^1 is compact.

The general notion of topological space is not sufficiently restrictive for most purposes. Therefore some additional axioms are almost always assumed—in particular the Hausdorff *separation axiom*. A topological space X satisfies the Hausdorff separation axiom, or is a Hausdorff space, if for every two distinct points x and x' of X there are open subsets U and V of X such that $x \in U$, $x' \in V$, and $U \cap V$ is empty. If X and Y are compact Hausdorff spaces and $f: X \to Y$ is a continuous map such that $f(x) = f(x')$ implies $x = x'$ and for every $y \in Y$ there exists $x \in X$ such that $f(x) = y$, then f is a homeomorphism.

All the topological spaces so far considered, the spheres, the euclidean spaces, and the subspaces or products of these spaces are Hausdorff spaces. In fact any subspace of a Hausdorff space, and any product of Hausdorff spaces, is again a Hausdorff space. Further if X is a topological space such that the topology is derived from a metric on X, then X is a Hausdorff space.

Manifolds. Though Hausdorff spaces form a much more interesting class of spaces than general topological spaces, the most important class of topological spaces, which is still much smaller, consists of the manifolds. An n-manifold is a Hausdorff space X such that the following conditions exist:

1. *For every point $x \in X$ there is an open subset U_x of X such that $x \in U_x$ and such that U_x is homeomorphic with an open subset of R^n.*
2. *There exists a metric ρ $X \times X \to R$ such that the topology on X is induced by the metric.*

In defining n-manifold one always assumes condition 1, which states that within short distances of some fixed point it seems as if one is in euclidean n-space. Having assumed condition 1, there are several other conditions which are sometimes assumed instead of condition 2. For example it is sometimes assumed that there exist compact subsets $X_1, X_2, \ldots, X_k, \ldots$ of X such that $X = \bigcup_{n=1}^{\infty} X_n$. Another frequent assumption is that there exists a countable number of points $x_1, x_2, \ldots, x_k, \ldots$ of X such that if U is any nonempty open subset of X, then there is some integer k such that $x_k \in U$ All these possible variations of condition 2 are essentially equivalent once condition 1 is assumed.

The spheres and the euclidean spaces are examples of manifolds. Further products of spheres and euclidean spaces are manifolds, for if X is an m-manifold and Y is an n-manifold, then $X \times Y$ is an $(m + n)$-manifold. This implies in particular that the torus $T = S^1 \times S^1$ is a 2-manifold.

A manifold is a topological space which is an n-manifold for some positive integer n. The dimension of this topological space is n. Moreover, the dimension of any open subset is also n.

Before proceeding further with the discussion of manifolds, it is necessary to introduce another general topological notion. A topological space X is connected if, whenever U and V are nonempty open subsets of X such that $X = U \cup V$, the set $U \cap V$ is nonempty. In other words X is connected if it cannot be expressed as the union of two disjoint nonempty open subsets.

[illegible]

[illegible] is a connected 1-manifold then either X is compact and is homeomorphic with S^1, or X is not compact and is homeomorphic with R^1. Thus from the point of view of topology these are just two connected 1-manifolds, namely the circle and the line.

The problem of classifying compact connected 2-manifolds has also been solved. These manifolds, also called *surfaces*, are classified by giving a list of standard surfaces, and then proving that any compact connected 2-manifold is homeomorphic to one and only one in the list. The list just mentioned is long and will not be given here. The 2-sphere, euclidean 2-space, and the torus are all examples of 2-manifolds. However, as mentioned previously, R^2 is not compact.

Little is known about the problem of classifying manifolds of dimension greater than 2. Considerable work has been done on the problem of classifying 3-manifolds, but the results are still far from satisfactory.

Homotopy theory. In recent years one of the most active branches of topology has been homotopy theory. If X and Y are topological spaces, then two continuous maps $f_0 . X \to Y$ and $f_1 : X \to Y$ are said to be homotopic if there exists a continuous map $F \cdot I \times X \to Y$ such that $F(0,x) = f_0(x)$ and $F(1,x) = f_1(x)$, where I is the subspace of the real numbers consisting of those numbers which are greater than or equal to 0 and less than or equal to 1 A map $f \cdot X \to Y$ is said to be a trivial map if $f(x) = f(x')$ for every pair of points x and x' belonging to X. A map $f_0 \cdot X \to Y$ is homotopically trivial if there exists a trivial map f_1 which is homotopic to f Intuitively two maps are homotopic if one can be smoothly deformed into the other. Thus a map is homotopically trivial if it can be deformed into a map which sends everything into a single point.

Suppose that X is a connected 2-manifold and every map $f \cdot S^1 \to X$ is homotopically trivial, then either X is compact and is homeomorphic with S^2, or X is not compact and is homeomorphic with R^2

Two topological spaces X and Y have the same homotopy type if there exist maps f $X \to Y$ and $g . Y \to X$ such that $f \cdot g$ $Y \to Y$ is homotopic to the identity map of Y and $g \cdot f$ $X \to X$ is homotopic to the identity map of X Consequently homeomorphic spaces always have the same homotopy type, but the converse is far from true. For example, for any positive integer n, the space R^n has the same homotopy type as the space consisting of a single point.

If X is a connected n-manifold, and for every positive integer q less than n every map f $S^q \to X$ is homotopically trivial, then either X is compact and has the same homotopy type as the sphere S^n, or X

is not compact and has the same homotopy type as R^n. The fact that at present the preceding result cannot be replaced by one saying that under the same conditions X is either homeomorphic with S^n or R^n unless n is less than or equal to 2, seems to be one of the main obstacles to proving classification theorems for manifolds of dimension greater than 2.

In the last fifty years the study of topology has changed considerably. Instead of attacking problems directly, algebraic invariants are attached to topological spaces, and then these invariants are studied. The main impetus to starting to work in this direction was given by the French mathematician Henri Poincaré about fifty or sixty years ago. The work of Poincaré was done chiefly in connection with the problem of classifying manifolds. He conjectured that if X is a compact connected 3-manifold and every ... $S^1 \to X$ is homotopically trivial, then X is ... mentioned ...

The study of topology by means ... invariants of topological spaces is in reality a comparatively new field of mathematics. Since most of these invariants are rather difficult to describe, no attempt to do so will be made here. The definition of just one set of these invariants, the homotopy groups will be outlined. These invariants are the easiest to describe, but among the most difficult to compute.

A topological space X is pathwise connected if for every pair of points x and x' belonging to X, there is a continuous map $f: I \to X$ such that $f(0) = x$ and $f(1) = x'$, in other words if one can draw an arc between any two points of the space. A manifold is connected if, and only if, it is pathwise connected.

Suppose that X and Y are topological spaces, $x_0 \in X$ and $y_0 \in Y$ are chosen and called base points. A map $f: X \to Y$ is said to preserve base points if $f(x_0) = y_0$. Two maps $f: X \to Y$ and $g: X \to Y$ which preserve base points are homotopic relative to the base points x_0 and y_0 if there exists a map $F: I \times X \to Y$ such that the following conditions obtain:

1. $F(0,x) = f(x)$ for any $x \in X$
2. $F(1,x) = g(x)$ for any $x \in X$
3. $F(t,x_0) = y_0$ for any $t \in I$

Now one says that two maps $f,g: X \to Y$ which preserve base point are equivalent if they are homotopic relative to the base points. The set of such equivalence classes of maps is denoted by $\pi((X,x_0),(Y,y_0))$. Such an equivalence class is called a homotopy class of maps.

For the sphere S^n choose a base point $e_n \in S^n$ once and for all. Let $\pi_n(X,x_0)$ be $\pi((S^n,e_n),(X,x_0))$ for every space X with base point x_0. Let $S^n \vee S^n$ be the space obtained by taking two copies of S^n and identifying the point e_n in one copy with the point e_n in the other copy. This space may be thought of as two tangent n-spheres. Define a map $\theta: S^n \to S^n \vee S^n$ by collapsing an equator of S^n through the base point to obtain two tangent spheres. Suppose that $f,g: S^n \to X$ are maps which preserve base points. Define $f \vee g: S^n \vee S^n \to X$ by mapping points of the first tangent sphere by means of f and of the second by means of g. This definition is legitimate since $f(e_n) = x_0 = g(e_n)$. Now $(f \vee g) \circ \theta: S^n \to X$ and preserves base point. One verifies that if f' is homotopic to f and g' is homotopic to g, then $(f' \vee g') \circ \theta$ is homotopic to $(f \vee g) \circ \theta$ where all homotopies are relative to the base points. If for any such map f, one denotes the homotopy class of f by $[f]$, then $[(f \vee g) \circ \theta]$ depends only on $[f]$ and $[g]$. Then mapping $\varphi: \pi_n(X,x_0) \times \pi_n(X,x_0) \to \pi_n(X,x_0)$ defined by

$$\varphi([f],[g]) = [(f \vee g) \circ \theta]$$

determines a group operation in the set $\pi_n(X,x_0)$ which is now called the n-dimensional homotopy group of X relative to the base point x_0.

If X is pathwise connected, then the group $\pi_n(X,x_0)$ is independent of the choice of base point x_0. Further if X and Y are pathwise connected spaces having the same homotopy type, then for any $x_0 \in X$ and any $y_0 \in Y$ the groups $\pi_n(X,x_0)$ and $\pi_n(Y,y_0)$ are isomorphic.

The group $\pi_1(X,x_0)$ is also called the fundamental group or Poincaré group of the space X based at the point x_0. This group was discovered and investigated b... ... were discovered ... opy type of a point, then $\pi_n(X,x_0)$ has a single element. Further it may be proved that the group $\pi_q(S^n,e_n)$ has a single element if q is less than n, but is isomorphic with the group of integers if $q = n$. This proves that S^n does not have the homotopy of a point.

Even though the groups $\pi_q(X,x_0)$ are abelian for $q > 1$, they are difficult to compute. It may be shown that if n is greater than 1, then the group $\pi_q(S^n,e_n)$ has more than one element for an infinite number of integers q.

[J C MO]

Bibliography: S. Lefschetz, *Introduction to Topology*, 1949; N. Steenrod and S. Eilenberg, *Foundations of Algebraic Topology*, 1952; R. L. Wilder, *Topology of Manifolds*, 1949.

Torbanite

A variety of coal that resembles a carbonaceous shale in outward appearance. It is fine-grained, brown to black, tough, and breaks with a conchoidal or subconchoidal fracture. The name torbanite is derived from the initial discovery site of the material in 1850 at Torbane Hill, Linlithgowshire, Scotland. Torbanite is synonymous with boghead coal and is related to cannel coal. Torbanite is derived from colonial algae identified with the modern species of *Botryococcus braunii* Kutz and its antecedent forms.

Major deposits of torbanite occur in Australia, Tasmania, New Zealand, Scotland, and South Africa. The South African deposit, which is in the Ermelo district of the Transvaal, yields from 20–100 gal of oil per ton on retorting. High-assay torbanite yields paraffinic oil, whereas low-assay material yields asphaltic oil. See COAL; SAPROPEL.

[I A B]

Torch

A gas mixing and burning tool that produces a hot flame for the welding or cutting of metal. The torch usually delivers acetylene and commercially pure oxygen producing a flame temperature of 5,000–6,000°F, sufficient to melt the metal locally. The torch thoroughly mixes the two gases and permits adjustment and regulation of the flame. Acetylene requires 2.5 times its volume of oxygen for complete combustion and, being an endothermic compound of carbon and hydrogen, can produce a higher flame temperature than other fuel gases. *See* ACETYLENE; WELDING AND CUTTING OF METALS.

Torches are of two types: low-pressure and high-pressure. In a low-pressure or injector torch, acetylene enters a mixing chamber where it meets a jet of high pressure oxygen (Fig. 1). The amount of acetylene drawn into the flame is controlled by velocity of this oxygen jet. In a high-pressure to both gases are delivered under pressure. The l developed at the work is controlled to some ex by gas pressure but principally by the size no or tip fitted to the torch. The larger the tip greater the required gas pressure. Small flames used with thin gage metals; large flames are ne sary for thick metal parts.

A welding torch mixes the fuel and gas intern and well ahead of the flame (Fig. 2). For cutti the torch delivers an additional jet of pure oxy to the center of the flame. The oxyacetylene fla produced by the internally mixed gases raises metal to its ignition temperature. The central o gen jet oxidizes the metal, the oxide being blo away by the velocity of the gas jet to leave a r row slit or kerf. In the case of iron, the oxides f

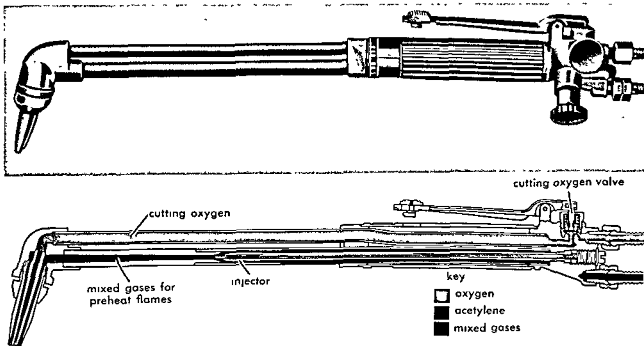


Fig. 1 Low-pressure injector cutting torch. (*Linde Co.*)

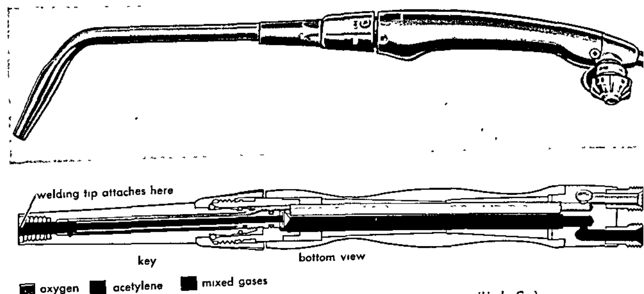


Fig. 2. Welding torch with cartridge mixer operates over wide pressure range. (*Linde Co.*)

t a lower temperature than the iron or steel so
at the oxides form, melt, and blow away before

ately cut in low-carbon steel by torches automati-
ally guided, such precision cutting being called
flame machining. [F H R.]

Tornado

An intense rotary storm of small diameter, the most violent of weather phenomena. Tornadoes always extend downward from the base of a convective-type cloud, generally the cumulonimbus of a large thunderstorm

Appearance ranges from a broad funnel with smallest diameter at the ground, to a narrow rope-like vortex, which may not reach the ground or may intermittently lift and dip. An ill-defined cloud of



Ground view photograph of tornado, June 20, 1957, at Fargo, North Dakota. (*Fargo Forum Photograph by C Gebert, Grand Prize Winner, 11th Annual Graflex Contest*)

dust or debris often surrounds the true tornado

structural damage and other evidence, probable wind speeds up to 300 mph or more have been calculated.

The visible funnel consists of cloud droplets condensed because of expansional cooling resulting from markedly lower (probably by 100–200 millibars) pressure in the vortex than in the surround-

distance upward within the accompanying cloud

Width of the path of destruction varies from a few yards to a mile, averaging 700 ft. Length of path ranges from very short up to 300 miles, averaging 4 or 5 miles. Movement is generally from southwest, but may be in any direction. Speed of movement averages 35 mph but is highly variable.

Damage in the millions of dollars, with loss of many lives, takes place occasionally when tornadoes strike heavily populated areas. Structural damage to buildings results in part from explosion when the atmospheric pressure outside is suddenly reduced, and partly from force of the extremely strong winds. Damage from explosion may be reduced by venting, or by prior opening of windows to allow rapid equalization of pressure inside and outside the building.

Tornadoes occur on all continents, but are common only in the United States and Australia. They have been observed in all states, being most frequent in Iowa, Kansas, Missouri, Illinois, and Oklahoma. Although tornadoes occur in all months, greatest seasonal frequency is in late spring and early summer. Most frequent occurrence is in the southern states in early spring; the locus of greatest activity shifts northward into the central states in summer. Occurrences are noted at all times of day, but there is a strong peak in incidence during mid-afternoon.

Requisite conditions for tornado formation are pronounced thermodynamic instability combined with sufficient amounts of water vapor to produce thunderstorms, along with the presence of strong winds in the upper troposphere. These conditions are most favored on the southeast sides of extratropical cyclones located east of the Rocky Mountains In the warm sector of the cyclone, southerly winds import warm, moist tropical air in lower levels northward from the Gulf of Mexico. At the same time, local cooling may take place higher up, as a cold low-pressure trough approaches from the west As a result of these processes combined with low-level solar heating, the required thermodynamic instability may be built up. Release of instability results either from frontal lifting, as at a cold front, or as a consequence of the general rising motions on the advancing side of the cyclone. *See* STORM; THUNDERSTORM.

Once thunderstorms have formed, often in the form of a squall line in advance of the cold front (*see* SQUALL), tornadoes may appear as parasites to the thunderstorms. With a single cyclone, families of as many as 10–30 tornadoes sometimes occur

No generally accepted theory of tornado mechanics has been formulated. Thermodynamic theories visualize the tornado as being a result of violent localized upward convection, with formation of a whirl from the compensating inrush of air in lower levels Mechanical theories suggest that the rotation is derived from interaction of currents having different directions and speeds, either at a given level or at different levels within or around the accompanying thunderstorms.

Tornadoes in the United States are mostly found on the south sides of the parent thunderstorms Heavy rain and hail often follow (but sometimes precede) passage of a tornado The heaviest rain is likely to fall a few miles north of the tornado track, and sometimes no rain falls along the track itself. Widespread thundersqualls are often observed outside the actual tornado path.

Accurate location of tornadoes by use of radar and radio direction-finding devices (frequent electrical discharges are characteristic) sometimes en-

Torch

A gas mixing and burning tool that produces a hot flame for the welding or cutting of metal. The torch usually delivers acetylene and commercially pure oxygen producing a flame temperature of 5,000–6,000°F, sufficient to melt the metal locally. The torch thoroughly mixes the two gases and permits adjustment and regulation of the flame. Acetylene requires 2.5 times its volume of oxygen for complete combustion and, being an endothermic compound of carbon and hydrogen, can produce a higher flame temperature than other fuel gases. *See* ACETYLENE; WELDING AND CUTTING OF METALS.

Torches are of two types: low-pressure and high-pressure. In a low-pressure or injector torch, acetylene enters a mixing chamber where it meets a jet of high pressure oxygen (Fig. 1). The amount of acetylene drawn into the flame is controlled by the velocity of this oxygen jet. In a high-pressure torch, both gases are delivered under pressure. The heat developed at the work is controlled to some extent by gas pressure but principally by the size nozzle or tip fitted to the torch. The larger the tip the greater the required gas pressure. Small flames are used with thin gage metals; large flames are necessary for thick metal parts.

A welding torch mixes the fuel and gas internally and well ahead of the flame (Fig. 2). For cutting, the torch delivers an additional jet of pure oxygen to the center of the flame. The oxyacetylene flame produced by the internally mixed gases raises the metal to its ignition temperature. The central oxygen jet oxidizes the metal, the oxide being blown away by the velocity of the gas jet to leave a narrow slit or kerf. In the case of iron, the oxides fuse

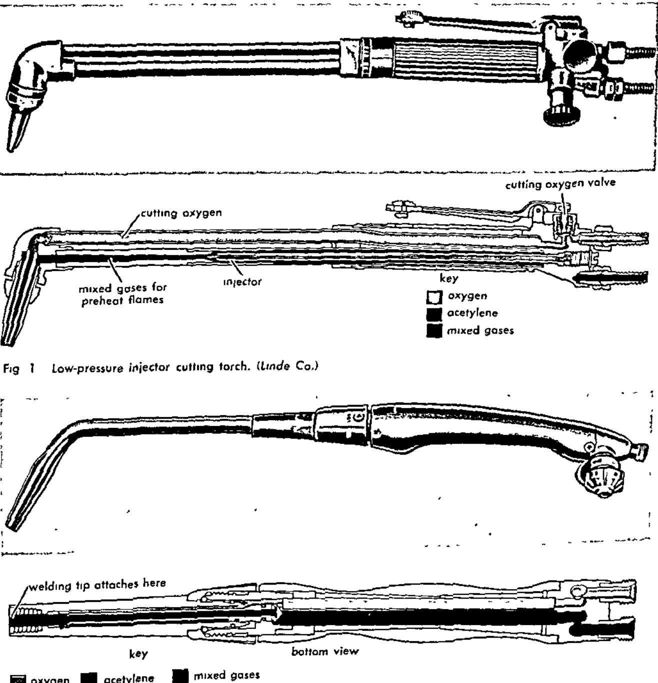


Fig 1 Low-pressure injector cutting torch. (*Linde Co.*)

Fig 2. Welding torch with cartridge mixer operates over wide pressure range. (*Linde Co.*)

t a lower temperature than the iron or steel so that the oxides form, melt, and blow away before the adjacent metal fuses. The temperature for the cutting action, once initiated, is maintained by the oxidization of the iron. Intricate shapes are accurately cut in low-carbon steel by torches automatically guided, such precision cutting being called flame machining. [F.H.R.]

Tornado

An intense rotary storm of small diameter, the most violent of weather phenomena. Tornadoes always extend downward from the base of a convective-type cloud, generally the cumulonimbus of a large thunderstorm.

Appearance ranges from a broad funnel with smallest diameter at the ground, to a narrow rope-like vortex, which may not reach the ground or may intermittently lift and dip. An ill-defined cloud of



Ground view photograph of tornado, June 20, 1957, at Fargo, North Dakota. (*Fargo Forum Photograph by C. Gebert, Grand Prize Winner, 11th Annual Graflex Contest*)

dust or debris often surrounds the true tornado cloud near the ground. In surface layers, air spirals inward toward the vortex, generally rotating in a counterclockwise sense, rising very rapidly in the core. From structural damage and other evidence, probable wind speeds up to 300 mph or more have been calculated.

The visible funnel consists of cloud droplets condensed because of expansional cooling resulting from markedly lower (probably by 100–200 millibars) pressure in the vortex than in the surroundings. Height of the visible funnel depends upon the cloud base, and may be 1000–10,000 ft; however, the tornado vortex probably extends a considerable distance upward within the accompanying cloud.

Width of the path of destruction varies from a few yards to a mile, averaging 700 ft. Length of path ranges from very short up to 300 miles, averaging 4 or 5 miles. Movement is generally from southwest, but may be in any direction. Speed of movement averages 35 mph but is highly variable.

Damage in the millions of dollars, with loss of many lives, takes place occasionally when tornadoes strike heavily populated areas. Structural damage to buildings results in part from explosion when the atmospheric pressure outside is suddenly reduced, and partly from force of the extremely [illegible] Damage from explosion may be re-

[illegible]

in Iowa, Kansas, Missouri, Illinois, and Oklahoma. Although tornadoes occur in all months, greatest seasonal frequency is in late spring and early summer. Most frequent occurrence is in the southern states in early spring; the locus of greatest activity shifts northward into the central states in summer. Occurrences are noted at all times of day, but there is a strong peak in incidence during mid-afternoon.

Requisite conditions for tornado formation are pronounced thermodynamic instability combined with sufficient amounts of water vapor to produce thunderstorms, along with the presence of strong winds in the upper troposphere. These conditions are most favored on the southeast sides of extratropical cyclones located east of the Rocky Mountains. In the warm sector of the cyclone, southerly winds import warm, moist tropical air in lower

[illegible]

west. As a result of these processes combined with low-level solar heating, the required thermodynamic instability may be built up. Release of instability results either from frontal lifting, as at a cold front, or as a consequence of the general rising motions on the advancing side of the cyclone. *See* STORM; THUNDERSTORM

Once thunderstorms have formed, often in the form of a squall line in advance of the cold front (*see* SQUALL), tornadoes may appear as parasites to the thunderstorms. With a single cyclone, families of as many as 10–30 tornadoes sometimes occur.

No generally accepted theory of tornado mechanics has been formulated. Thermodynamic theories visualize the tornado as being a result of violent localized upward convection, with formation of a whirl from the compensating inrush of air in lower levels. Mechanical theories suggest that the rotation is derived from interaction of currents having different directions and speeds, either at a given level or at different levels within or around the accompanying thunderstorms.

Tornadoes in the United States are mostly found on the south sides of the parent thunderstorms. Heavy rain and hail often follow (but sometimes precede) passage of a tornado. The heaviest rain is likely to fall a few miles north of the tornado track, and sometimes no rain falls along the track itself. Widespread thundersqualls are often observed outside the actual tornado path.

Accurate location of tornadoes by use of radar and radio direction-finding devices (frequent electrical discharges are characteristic) sometimes en-

ables useful short-period prediction of likely future movements. [C.W.N.]

Bibliography: T. F. Malone (ed.), *Compendium of Meteorology*, 1951; S. D. Flora, *Tornadoes of the United States*, rev. ed., 1954; S. Petterssen, *Weather Analysis and Forecasting*, 2d ed. 1956.

Torque

The product of a force and its perpendicular distance to a point of turning, also called the moment of the force. Torque produces torsion and tends to produce rotation. Torque arises from a force or forces acting tangentially to a cylinder, or from any force or force system acting about a point. A couple, consisting of two equal, parallel, and oppositely directed forces produces a torque or moment about the central point. A prime mover such as a turbine exerts a twisting effort on its output shaft, measured as torque. In structures, torque appears as the sum of moments of torsional shear forces acting on a transverse section of a shaft or beam. See COUPLE; TORSION [N.S.F.]

Torque converter

A device for changing the torque-speed ratio or mechanical advantage between an input shaft and an output shaft. A pair of gears is a mechanical torque converter (*see* GEAR DRIVE). An hydraulic torque converter, with which this article deals, is an automatically and continuously variable torque converter, in contrast to a gear shift whose torque ratio is changed in steps by an external control. *See* TRANSMISSION, AUTOMOTIVE.

Converter characteristics. A mechanical torque converter transmits power with only incidental losses; thus, the power, which is the product of torque T and rotational speed N, at input I is substantially equal to the power at output O of a mechanical torque converter, or $T_I N_I = kT_O N_O$, where k is the efficiency of the gear train This equal-power characteristic is in contrast to that of a fluid coupling in which input and output torques are equal during steady-state operation *See* FLUID COUPLING.

In an hydraulic torque converter, efficiency depends intimately on the angles at which the fluid enters and leaves the blades of the several parts. Because these angles change appreciably over the operating range, k varies, being, by definition, zero when the output is stalled, although output torque at stall may be three times engine torque for a single-stage converter and five times engine torque for a three-stage converter. Depending on its input absorption characteristics, the hydraulic torque converter tends to pull down the engine speed toward the speed at which the engine develops maximum torque when the load pulls down the converter output speed toward stall.

Converter power efficiency is highest (80–90%) at a design speed, usually 40–80% of maximum engine speed, and falls toward zero as shaft speed approaches engine speed. Because of this characteristic, the mode of operation may be modified to change from torque conversion to simple fluid coupling or to direct mechanical drive at high speed.

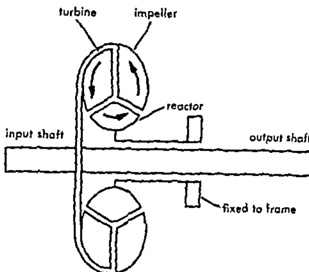


Fig. 1. Elementary hydraulic torque converter.

Hydraulic action. These characteristics are achieved by the exchange of momentum between the solid parts of the converter and the fluid (Fig. 1) A vaned impeller on the input shaft pumps the fluid from near the axis of rotation to the outer rim. Fluid momentum increases because of the greater radius and the influence of the vanes. The high-energy fluid leaves the impeller and impinges on the

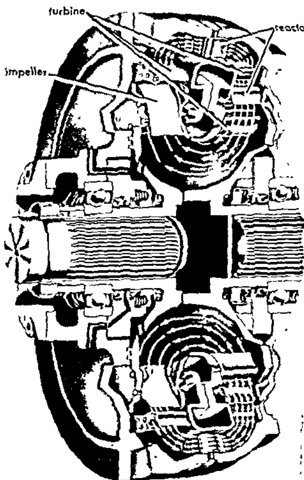


Fig. 2. Three-stage converter showing simplified fluid flow around torus. (Twin Disc Clutch Co.)

blad(
the t
The ...
reactor. The reactor blades are fixed to the frame, they deflect the fluid flow and redirect it into the impeller. This change in flow direction produced by the stationary reactor is equivalent to an increasing change in momentum which adds to the momentum imparted by the impeller to give a torque increase at the output of the converter (Fig. 2)

In a typical converter, as the output shaft comes up to the speed of the input shaft, efficiency de-

the reaction member to give a four-element polyphase converter gives even more uniform efficiency. [H J WI]

Bibliography: S. J. Berard, E O Waters, and C. W. Phelps, *Principles of Machine Design*, 1955

Torricelli's theorem

The speed of efflux of a liquid from an opening in a reservoir equals the speed the liquid would acquire if allowed to fall from rest from the surface of the reservoir to the opening

Torricelli, a student of Galileo, observed this relationship in 1643 In equation form $v^2 = 2gh$, in which v is the speed of efflux, h the head (or elevation difference between reservoir surface and centerline of opening if in a vertical plane), and g the acceleration due to gravity. (The equation is the same as that for a solid particle dropped a distance h in a vacuum) The relationship can be derived from the energy equation for flow along a streamline, if energy losses are neglected

An orifice (opening in the wall or bottom of a reservoir) is used as a flow measuring device From Torricelli's theorem, by solving for v and multiplying by the flow area, an expression for discharge Q, in volume per unit time, is obtained In equation form $Q = C_d A\sqrt{2gh}$, in which A is the area of opening and C_d is a dimensionless coefficient, determined experimentally, that corrects for contraction of the jet as it leaves the orifice and for energy loss due to viscosity. When h is measured Q may be determined from the formula. *See* FLOW MEASUREMENT [V L S.]

Torsion

A straining action produced by terminal couples that act normal to the axis of a member Torsion is identified by a twisting deformation.

In practice, torsion is often accompanied by bending or axial thrust as in the case of line shafting driving gears or pulleys, or propeller shafts for ship propulsion. Other important examples include springs and machine mechanisms usually having circular sections, either solid or tubular Members with noncircular sections are of interest in special applications such as structural members subjected to unsymmetrical bending loads that twist and buckle beams.

When subjected only to torque, the member is in pure torsion, which produces *pure shear stresses* (*see* SHEAR). The shear properties of materials are determined by a *torsion test*.

Cylindrical bars. The twist of a bar due to *torque* can be visualized as the accumulated rotational displacements of imaginary disks cut by transverse sections on which tangential forces operate. Shearing forces vary across the section and together furnish the internal resisting torque.

Torsional angle, designated θ, is the total relative rotation of the ends of a straight cylindrical bar of length L, when subjected to torque.

Helical angle, designated ϕ, is the angular displacement of a *longitudinal element*, originally straight on the surface of the untwisted bar, which becomes helical after twisting (Fig. 1). Angle ϕ is the shear strain. For small twist, torsional and helical angles are related by geometry $\phi = (R/L)\theta$, where R is the radius of the bar.

Elastic shear stress. Within the elastic limit, shear stress S_s is found by Hooke's law $S_s/\phi = E_s$, and is expressed in terms of the torsional angle as $S_s = (R/L)E_s\theta$, where E_s is the modulus of rigidity *See* HOOKE'S LAW.

The shear stress varies linearly across the section, being maximum at the surface and zero at the center For a circular section the maximum shear stress acting perpendicular to the radius at the extreme distance $R = D/2$ from the neutral axis is $S_{max} = 16T/\pi D^3$ where T is the externally applied twisting moment.

Tangential shear stresses on the section are accompanied by longitudinal shear stresses along the bar These complimentary stresses induce tensile and compressive stresses, equal to the shear intensity, at 45° to the shear stresses The longitudinal stresses are important in laminated materials, wood, or metals with seams Brittle materials, low in tensile strength, fracture on a 45° helicoidal surface; ductile materials fracture on transverse sections after large twist

Resisting torque equal to the applied torque is the moment of the elementary internal shear forces about the neutral axis expressed in terms of the sectional dimensions and the stresses. A general expression for resisting torque is $T = S_{max}J/R$, where J is the polar moment of inertia of the section This relation is applicable to both solid and

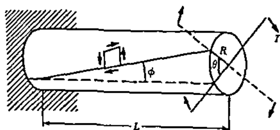


Fig 1 Cylindrical bar in torsion.

tubular circular sections which are differentiated by J. In terms of torque T, torsional angle θ is TL/E_sJ. Torsional angle per unit of length is a measure of torsional stiffness, which may limit the required dimensions of a shaft. In power transmission, the torque associated with horsepower is found from HP = $TN/63000$, where T is expressed in inch-pounds of moment and N is the rotation of the shaft in revolutions per minute.

Inelastic behavior in torsion. Strains exceeding the elastic limit are not completely recoverable after unloading and the behavior is inelastic. Torsional strains vary linearly from the center of the bar during both elastic and plastic deformation, and the corresponding shear stresses reflect the stress-strain curve for the material (Fig. 2). After the extreme element reaches the yield point, continued twisting produces inelastic strains at increasing distances from the surface while the stress remains constant. When the action is fully plastic, the stresses are constant, equal to the yield point over the entire section. The fully plastic resisting torque is

$$T_p = \frac{4}{3}\frac{S_{yp}J}{R}$$

which is 1.33 times that required to just produce surface yielding. Torsional resistance increases due to strain hardening but is of interest only where large deformation can be tolerated. Elastic analysis is applicable to designs where permanent deformation must be avoided and where endurance (fatigue) properties limit the stresses.

Thin-walled tubes. Thin tubular members find application particularly in aircraft. Shear stresses are assumed uniform over the wall thickness, when a thin-walled tube of any shape is subjected to torque at the ends Shear force q per unit length of perimeter is constant.

Shear flow is the constant shear force q acting along the median line of the wall and is equal to the product of shear stress S times thickness t at any point, thus $q = St$ is constant. The concept of flow is drawn from the similarity of the expression for constant shear force with the constant quantity Q of a liquid passing variable sections of a channel having area A and velocity V, $Q = AV$. Resisting torque T is the summation of moments of

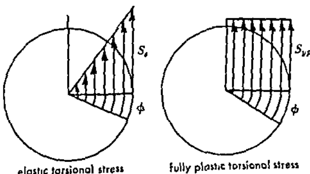


Fig. 2. Stress distribution.

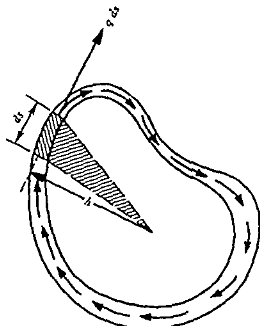


Fig. 3. Shear flow

shear forces on elementary lengths ds of the wall perimeter about the center of rotation $T = 2Aq$, where A is the area enclosed by the center line of the tube wall (Fig 3) The stress at any point where thickness is t is $S = q/t = T/2At$. The torsional angle produced by applied torque T is found from

$$\theta = \frac{TL}{4A^2E}\int_0^S \frac{ds}{t}$$

where S is the length of the perimeter and t is the variable thickness. For constant thickness, $\theta = TLS/4A^2Et$ where S is peripheral length of the center line.

Solid noncircular sections. When a solid member with noncircular section is twisted, the sections become warped and the stresses do not vary linearly as in the case of circular sections. Evaluation of stresses and torsional twist requires the rigorous procedures of the theory of elasticity. If a grid is scribed on the surface of a square or rectangular bar and the bar is twisted, distortions of the grid indicate that maximum shear stress is at a boundary nearest the center. Contrary to theory applicable to circular sections, the stress is zero at the corners, which are the most remote elements. The location of maximum stress is indicated by points of initial plastic yielding as shown by the macrographs of a square and a round bar (Fig. 4). Sections were etched after yielding, thus differentiating the darker plastic zones. Formulas for maximum shear stress and torsional angle for common noncircular sections are presented in Fig. 5

Helical springs subjected to axial loads involve all four possible straining actions: direct stress, transverse shear, bending, and torsional shear. For small obliquity of the coils, as in close-coiled

brane analogy. The analogy presented by Ludwig Prandtl (1903) is based on the similarity of the equilibrium equation for a membrane with pressure on one side and the differential equation for torsional stresses

In application, a thin membrane such as a soap film is placed over an opening in a plate, having the same geometrical shape as the section under investigation. Slight air pressure on one side deflects the film, and micrometer measurements determine the contours of equal deflection. The slope at any point and the volume enclosed by the deflected

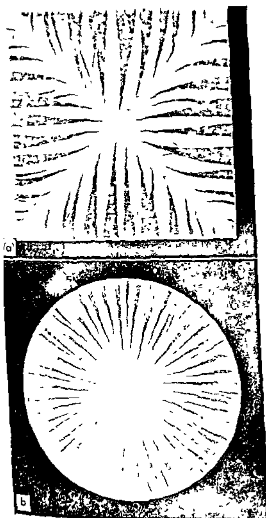


Fig 4 Plastic strain in torsion (a) Square bar (b) Round bar.

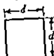
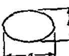
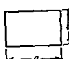
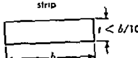
section	shear stress S_s and torsional angle θ
square 	$S_s = T/0.208d^3$ $\theta = 7.11TL/d^4E_s$
ellipse 	$S_s = 16T/\pi ab^2$ $\theta = 16TL(a^2 + b^2)/\pi E_s a^3b^3$
rectangle 	$S_s = (3a + 1.8b)T/a^2b^2$ $\theta = 10TL(a^2 + b^2)/3E_s a^3b^3$
strip 	$S_s = 3T/bt^2$ $\theta = 3TL/bt^3E_s$
equilateral triangle 	$S_s = 20T/a^3$ $\theta = 80TL/\sqrt{3}\,a^4E_s$

Fig 5. Shear stress and torsional angle for common noncircular sections

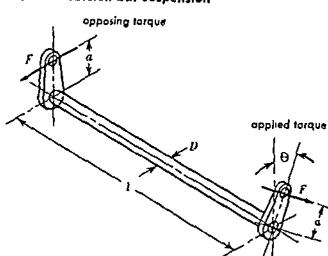
membrane can be found from these measurements. If a bar having this section is twisted, the torsional shearing stress at any point is proportional to the slope of the membrane, the stress direction is tangent to the contour, and the torque is proportional to the volume enclosed by the deflected membrane.

The method is a valuable qualitative aid in locating points of maximum stress by visualizing or observing points of maximum slope of the deflected film The high stress at a reentrant corner such as at a fillet of a structural angle or channel section is indicated by a steep slope of the film [W J KR]

Bibliography: E Murphy, *Advanced Mechanics of Materials*, 1946; S Timoshenko and J. N. Goodier, *Theory of Elasticity*, 2d ed , 1951.

Torsion bar

A spring flexed by twisting about its axis Design of a torsion bar spring is primarily based on the relationships between the torque applied in twisting the spring, the angle through which the torsion bar twists, and the physical dimensions and material (modulus of elasticity in shear) from which the torsion bar is made The drawing shows the elements of a simple torsion bar and the impor-



Torsion bar.

tant dimensions involved in its design. The equation relating these dimensions is

$$\theta = \frac{32Fal}{\pi D^4 G}$$

in which θ is angle of twist in radians, F is force in pounds, a is radius arm of force in inches, l is length of torsion bar in inches, D is diameter of torsion bar in inches, and G is modulus of elasticity in shear in pounds per square inch.

If the deflection or twist of the spring θ is large, force F must change direction if a is to remain constant. For this reason the equation is frequently written

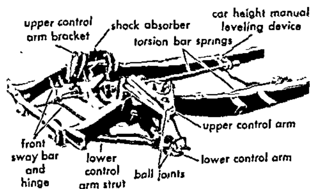
$$\theta = \frac{32\tau l}{\pi D^4 G}$$

in which τ is the torque in inch lb.

Torsion bar springs are found in the spring suspension of truck and passenger car wheels, in production machines where space limitations are critical, and in high-speed mechanisms where inertia forces must be minimized *See* SPRING (MECHANICAL). [L.S.L.]

Torsion bar suspension

A resilient support in which the spring is a bar or rod stressed in torsion about its long axis. The neutral or normal rest position of an automobile with



Torsion bar supports lower control arm of front wheel suspension. (*Chrysler Corp.*)

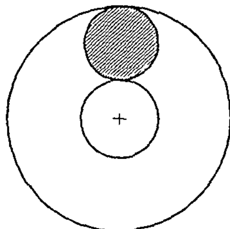
torsion bar suspension is easily changed aut matically or manually by adjusting the torque o the bar (as illustrated). The bar is rigidly fixed the car frame at the rear end and is free to rota at its front end near which the lower wheel contr arm is splined to the bar. A torsion bar provid stiffness in the direction of the bar. This stiffne aids steering and directional stability. [N.M

Tortoise

A name applied somewhat indiscriminately to var *ous turtles, notably the dry land forms, but wit* out any clear-cut definition. The name is fallin into disuse. *See* CHELONIA; TURTLE. [J.D.B.

Torus

A surface obtained by rotating a circle about a lin that lies in its plane, but which has no points i common with it. It is a 2-dimensional manifol of genus 1 and connectivity 3 (*see* MANIFOLD The equations $x = u \cos v$, $y = u \sin v$, $z =$ $[r^2 - (u - b)^2]^{1/2}$, $b > r > 0$, represent the uppe


A torus.

half of the torus obtained by rotating about the z axis a circle of radius r whose center is the point $(b,0,0)$. The parameter u represents the distance of a point P of the torus from the z axis, and v is the angle of rotation According to whether $b < u \leqq b + a$ or $b - a \leqq u < b$ or $u = b$, the corresponding point P is elliptic, hyperbolic, or parabolic, respectively, and the Gauss curvature of the surface at P is positive, negative, or zero (*see* GEOMETRY, DIFFERENTIAL). *See also* SURFACE AND SOLID OF REVOLUTION; TOPOLOGY [L.M BL.]

Touch

A term for both the generic designation for general bodily feeling and, more narrowly, the array of skin sensations ranging from contact to dull pressure and including light touch, granular pressure, and a host of other cutaneous patterns for which specific names have not entered the English language (*see* CUTANEOUS SENSATION; SOMESTHESIS) The system of pressure sensitivity of the skin yields a wealth of feeling patterns which differ among themselves along the spatial, temporal, and intensi-

tive dimensions. They have also been held to vary within a brightness-dullness continuum, by analogy with visual sensations. For certain of the patterns which occur repeatedly and familiarly, we have terms such as *lively contact*, *tickle*, *vibration*, and *deep pressure*. Others, though recognizable when felt, are nameless and thus impossible of description

Pressure thresholds. The normal, or adequate, stimulus for the arousal of pressure sensations is tension within the cutaneous tissues. Simple me-

over most of the finger, there is no feeling of pressure except for the ring at the surface of the liquid where there is an abrupt transition from no pressure to pressure. Differential pressure, a gradient, must be supplied before there is the necessary shearing force to create tissue tension. For small stimulators, hairs or fine needles pressed gently into the skin of the arm, the amount of tension that will just produce a pressure sensation is roughly 1 0 g/mm. About the same value is found if a thread is cemented to the skin and tension is brought about by an upward pull No single threshold figure is representative, however, for pressure sensitivity proves to vary with skin locus, the speed (or perhaps acceleration) with which the stimulus tip is applied, and especially the size of the stimulator. Thresholds, expressed in tension units (g/mm), are far from constant for stimulators much larger than hairs.

One highly efficient method for evoking touch sensations is to take advantage of the fact that hairs projecting from the skin, their follicles firmly set in cutaneous tissues beneath the surface act as levers of the second class when their distal ends are moved. It has been shown experimentally that

tiate a sensory system, it is actually about 5,000,000,000 times as much as the amount required to get the retina of the eye or the auditory nerve endings of the cochlea into minimal action

Since pressure sensitivity, like that for pain and temperature, is distributed throughout the skin as minute spots, another way to measure it is to ascertain the number of loci within a fixed area that will respond when a systematic exploration is made. If a horsehair of variable length and stiffness is employed in a series of successive explorations of the same skin area it will be found that the proportion of spots yielding pressure sensations will increase with stimulus intensity in an interesting way

Pressure adaptation. If a steady mechanical pressure is exerted against the skin it will be found that the intensity of the resulting pressure sensation will gradually decline and eventually disappear entirely. However, care must be taken to insure that extraneous forces are not allowed to alter

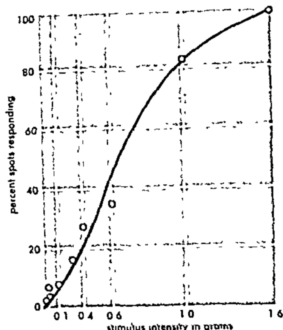


Fig. 1. Responsiveness of skin to pressure stimuli of graded strength. (E. G. Boring, H. S. Langfeld, and H. P Weld, *Foundations of Psychology*, Wiley, 1948)

the essential unvarying relation between the stimulator and the tissue on which it is acting. This phenomenon of adaptation is a universal one in the realm of sensation. Given steady stimulation, sensation fades

Whereas there is no reason to believe that the pressure receptors do not operate in accordance with the general principle, there is good evidence to indicate that there may be an additional mechanism hastening or supplementing the adaptation process. A weight placed on the skin surface continues to move downward, "sinking" into the skin for a surprisingly long time The subcutaneous tissues are somewhat compressible and take up the force of the stimulus, meanwhile allowing changing tensions (supraliminal stimulation) at receptive nerve endings It has been shown that sensation fades out just about at the time the falling weight comes to rest The implication is that tissue movement is the effective stimulus and that stimulus failure accounts for complete adaptation. This view is further reinforced by the observation that sudden removal of the weight rearouses pressure sensations, which continue in force until the cutaneous tissues once more attain their original disposition.

Spatial discrimination by touch. Vision and touch are the only two senses that have appreciable anatomical extension. Of the two, the eye far outstrips the skin in the capacity to gather spatial information The fact that light travels in straight lines is basic, of course, but it is also the case that the retina of the eye possesses a resolving power, by virtue of its fine microstructure, that is not even approximated by the skin. Some spatial discrimination is inherent in touch, however, and it is not difficult to find evidence of it.

One way to assess the skin's space-resolving ability is to ascertain the accuracy with which stimulation of its surface may be localized. If an observer is touched with a blunt stylus, vision being excluded, and he attempts to reproduce the feel, by touching himself with another stimulator or by marking the judged location on a map of the skin, it will be found that characteristic errors occur. On the lips or finger tips the average error may be no more than a millimeter. The volar forearm gives an error about eight times as large, and it is doubled again when the thigh is the site of stimulation.

Another way to measure tactual acuity is to determine the ability to resolve two simultaneous touches. This may be done with compass points—or better, a two-point esthesiometer, which precludes thermal stimulation—the distance apart of the contactors being progressively narrowed until the two feel like one. The entire body has been charted for the two-point threshold, the minimal separation of the esthesiometer points perceived as two contacts. It proves to vary widely, from tongue to small of the back, and to parallel the absolute error of localization, being about three to four times the size of the latter.

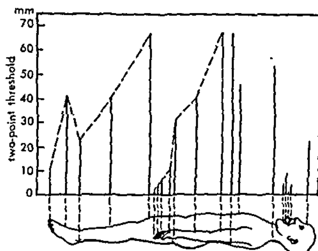


Fig. 2. Bodily variations in the two-point threshold (J F Fulton, ed., *Howell's Textbook of Physiology*, Saunders, 1947)

Perception of vibration. If a rapid succession of tiny impacts, such as that supplied by the base of a vibrating tuning fork, is applied to the skin, there is felt a continuous "whirring," a lively, sustained cutaneous pattern of somewhat indefinite localization. The magnitude of the impacts necessary to reach threshold depends on a number of factors: (1) bodily locus—vibration having an amplitude of less than 1 micron (μ) may be appreciated by the finger tip, while many times this amount are needed on the arm or leg; (2) frequency of vibration—a minimum frequency of about 16 cycles per sec (cps or ~) must be supplied if continuity is to be experienced, and a frequency of about 250 cps is most efficient, that is, yields the lowest thresholds; (3) the size of the skin area stimulated—in general and within limits, the larger the contactors, the smaller the amplitude required to reach threshold; (4) skin temperature—it has been shown that vibratory sensitivity increases, then declines, as the cutaneous area receiving it is warmed. Conversely, cooling reduces sensitivity.

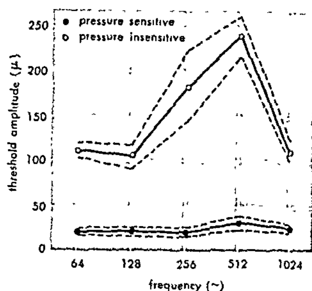


Fig. 3. Vibratory thresholds of two populations of cutaneous spots. Averages are represented. Reliability of the measures is indicated by the dotted lines, standard errors of the mean having been plotted. (F. A Geldard, The perception of mechanical vibration, *J Gen. Psychol*, 22:286, 1940)

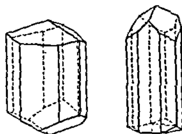
It has not always been apparent that vibratory sensitivity is handled by the same mechanism as that responsible for pressure sensations. There have been claims for a separate "vibratory sense" with its own machinery of reception and report. The case for this cutaneous pattern as being simply "pressure in movement" is strengthened by the finding that vibratory sensitivity is distributed in the skin in a punctiform manner and that, moreover, spots highly sensitive to mechanical pressure have low vibratory thresholds while relatively insensitive ones have high thresholds [F A G]

Bibliography: E G Boring, *Sensation and Perception in the History of Experimental Psychology*, 1942. J F. Fulton (ed.), *Howell's Textbook of Physiology*, 17th ed., 1955.

Tourmaline

A mineral cyclosilicate with complex chemical composition, long known for its use as a gem stone. *See* GEM; SILICATE MINERALS

Tourmaline crystallizes in the ditrigonal-pyramidal class of the hexagonal system in prismatic crystals with the trigonal prism dominant A combination of this prism with a hexagonal prism causes vertical striations and a tendency for the faces to round into each other, giving the crystals a cross section resembling a spherical triangle. The vertical axis is polar; thus different forms are found at the opposite ends Because of this polar-



Doubly terminated crystals of mineral tourmaline, showing different forms at opposite ends. (From C. S. *Hurlbut, Jr., Dana's Manual of Mineralogy, 16th ed., Wiley, 1952*)

ity, tourmaline is piezoelectric; that is, if pressure is exerted at the ends of the polar axis, one end becomes positively charged and the other end negatively charged. It is also pyroelectric, with the electrical charges developed at the ends of the polar axis on a change in temperature. Because of its piezoelectric property, tourmaline is manufactured into gages to measure transient pressures. Plates, cut at right angles to the principal axis, are coated with electrodes from which wires lead to a recording device. The voltage recorded is proportional to the pressure exerted on the plate. Such gages are used to measure the pressures of atomic explosions. See PIEZOELECTRICITY, PYROELECTRICITY.

The hardness of tourmaline is 7½ on Mohs scale and the specific gravity 3.0–3.25. The luster is vitreous to resinous and the color, depending on the chemical composition, is variable. Iron-rich tourmaline (shorlite), the most common variety, is black. The magnesium variety is brown. Lithium renders the mineral lighter-colored in various [illegible] (rubellite), yellow, green, blue (in- [illegible] stones. Several colors may be present [illegible] crystal, arranged in layers across the length or in concentric envelopes around the crystal. Some dark varieties are strongly dichoric. The chemical composition is represented by the general formula $XY_3Al_6(BO_3)_3Si_6O_{18}(OH)_4$, where X is Na,Ca and Y is Al, Fe^{3+}, Li,Mg.

Tourmaline is found as an accessory mineral in igneous and metamorphic rocks, but its most characteristic occurrence is in granite pegmatites. Here the black variety is most common but the light-colored varieties may be present, firmly embedded in the other pegmatite minerals or in cavities known as pockets. Most gem material occurs in this latter form. Noted localities for gem tourmaline are in Minas Gerais, Brazil; Ural Mountains; Madagascar, and, in the United States, Paris and Auburn, Maine, Haddam Neck, Connecticut; Mesa Grande and Pala, California. [C.S.HU]